THE UNIVERSITY *of* EDINBURGH

Mechanism and function of complement factor H

Nicola McIntosh

THE UNIVERSITY
of EDINBURGH

University of Edinburgh

Thesis submitted for the degree of Doctor of Philosophy

August 2013

Abstract

Factor H (FH) is a 155-kDa plasma protein that regulates the alternative pathway of the complement system. Its 20 CCP modules, of 51-62 amino acid residues each, are linked by short stretches (“linkers’) of three to eight residues. We set out to test the hypothesis that long linkers towards the middle of FH play a role in ensuring that its architecture allows binding sites near its N- and C-termini to engage cooperatively with the main target, C3b, which is the key complement pathway-triggering product of C3 cleavage.

In initial work, site-directed mutagenesis was used to test whether two mutations, R53H and R78G, located within CCP 1 and linked to the kidney disease atypical hemolytic uremic syndrome, are functionally deficient. Mutant versions and a native-sequence version of CCPs 1-4 of FH (i.e. FH 1-4) were tested for their ability to act as a cofactor for the FI-mediated cleavage of C3b, and accelerate the decay of the C3 convertase. It was shown that FH 1-4 R53H binds normally to C3b but has no regulatory activity while FH 1-4 R78G binds very poorly and is also deficient in co-factor and decay-accelerating activities.

In subsequent work, mutagenesis was used to make the eight-residue CCPs 12-13 linker shorter (SL), or more flexible through introduction of glycine residues (3xGLY), within recombinant (r) module pair FH 12-13, and in rFH 10-15 and rFH 8-15 as well as full-length rFH. NMR showed CCPs 12 and 13 remain intact following mutation of the linker but (in FH 12-13) are more flexibly mutually disposed, as expected. SAXS indicated that both FH 10-15 SL and FH 10-15 3xGLY nonetheless have similar compact structures to native sequence (WT) FH 10-15. On the other hand, FH linker mutants interact with C3b (according to surface plasmon resonance) somewhat less well than WT FH and in the case of FH SL, affinity is similar to that of FH 19-20, i.e. there is no evidence that both C3b-binding sites in this mutant bind to the target simultaneously. Nonetheless, the bacterial protein PspCN boosts binding of linker mutants to C3b by a similar factor (three-to-fivefold) to that observed for FH WT. Thus, while interactions between non-sequential CCPs

are important for FH architecture, a bend at the 12-13 linker is needed for full-length FH to adopt a fully biological activity confirmation.

The use of EPR for structural studies of rFH and its mutants was explored. Free cysteines were engineered in so they could have spin labels site-specifically attached. Alternatively, a recognition site for transglutaminase was introduced so a spin label could be incorporated. These strategies were applied to rFH 12-13 and rFH 10-15 as a prelude to studies of full-length FH. Several suitably engineered proteins were prepared but only one paramagnetically labeled sample (of FH 12-13) made it for EPR; this yielded results commensurate with the NMR-derived structure.Taken together, these promising data lay the groundwork for a future, potentially very insightful, combined mutagenesis and EPR study of FH architecture and its role in complement activation.

Acknowledgements

Firstly I would like to thank my supervisor, Prof Paul Barlow, for the opportunity he gave me and for his help, support and patience.

I owe a huge thank you to all the past and present members of Lab 120 who have helped at different stages during this PhD.

Thanks to John White for his friendliness and interest during the fermentation process. I would also like to express my gratitude to Dr David Norman and Dr Haydyn Mertens for their advice and recording of the data.

A special thanks to Eliza and Haris for keeping me sane during the PhD process and for all the fun away from the lab.

To my friends and their crazy kids for understanding that although I would have liked to have joined you it was not always possible.

Lastly to my family who were supportive and understanding when times got hard. Finally to Graeme, it was not always easy but we made it and now on to the next challenge.

Declaration

Unless stated in the text, the work described in this thesis is my own work and has not been submitted for in whole or in part for a degree or other qualification at this or any other university.

Nicola McIntosh

Abbreviations

1D	one-dimensional
2D	two-dimensional
α'NT	α-chain N-terminal residues
ABC	ammonium bicarbonate
ACN	acetonitrile
aHUS	Atypical hemolytic uremic syndrome
AMD	age-related macular degeneration
ANA	anaphylatoxin domain
AOX	alcohol oxidase
AP	alternative pathway
AUC	analytical ultracentrifugation
BMG	buffered minimal glycerol
BMGY	buffered glycerol-complex medium
BMMY	buffered methanol-complex
BSA	bovine serum albumin
C4BP	C4b-binding protein
CA	Cofactor activity
CCP	complement control protein
CIP	Calf Intestinal Alkaline Phosphatase
CP	classical pathway
CR1	complement receptor type 1
CW	continuous wave
Da	daltons
DAA	decay accelerating activity
DAF	decay accelerating factor
DCM	dichloromethane
DDD	dense deposit disease
DEER	double electron-electron resonance
DMSO	dimethyl sulfoxide
DMF	dimethylformamide

DNA	deoxyribonucleic acid
dNTPs	deoxynucleosides
DO	dissolved oxygen
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EM	transmission electron microscopy
EOM	ensemble optimization method
EPR	electron paramagnetic resonance
FB	factor B
FD	factor D
FH	factor H
FHL-1	factor H-like protein 1
FHR	factor H-related proteins
FI	factor I
Fmoc	Fluoroenylmethyloxycarbonyl chloride
FRET	förster resonance energy transfer
FT–ICR	Fourier-transform ion-cyclotron resonance
GAGs	glycosaminoglycan
GBM	glomerular basement membrane
HSQC	heteronuclear single quantum coherence
K_D	equilibrium dissociation constant
LB	Luria Bertani
LC-MS	liquid chromatography mass spectrometry
LNK	linker domain
LP	lectin pathway
MAC	membrane attack complex
MASP	MBL-associated serine proteases
MBL	mannose-binding lectin
MCP	membrane cofactor protein
MG	macroglobulin domains
MPGN II	Membrano-proliferative glomerulonephritis type II
MTSSL	1-oxyl-2,2,5,5–tetramethylpyrrol–3-yl-methyl methanethiosulfonate

NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PCR	polymerase chain reaction
PspC	pneumococcal surface protein C
PELDOR	pulse electron-electron double resonance
PMSF	phenylmethylsulfonyl fluoride
RCA	regulator of complement activation
RFU	relative fluorescence unit
RPE	retinal pigment epithelium
RU	response unit
SAXS	small-angle X-ray scattering
SDSL	site-directed spin labeling
SDS-PAGE	sodium dodecylsulfate-polyacrylamide gel electrophoresis
SOC	super optimal catabolite suppressor
SPR	surface plasmon resonance
TAE	Tris-acetate-EDTA
TEA	triethanolamine
TED	thioester domain
TGase	transglutaminase
THF	tetrahydrofuran
v/v	volume/volume ratio
w/v	weight/volume ratio
YPD	yeast-peptone-dextrose
YPDS	yeast-peptone-dextrose-sorbitol

Table of contents

List of figures

List of tables

Chapter 1 Introduction

1.1 Complement System

1.1.1 Historical perspective

Initial reports of complement were published as early as the 1890s by Buchner and Bordet who independently described a heat-labile component in serum that aided or "complemented" the killing of bacteria by antibodies [1]. Later in the same decade, the term "complement" was introduced by Paul Ehrlich. For many years it was thought that the role of complement was limited to the innate immune response, and to being an effector arm of antibody-mediated immunity. More recently it was recognized that complement also augments adaptive immunity by influencing both B-cells and T-cells [2, 3]. Furthermore, it is now realized that the complement system contributes to homeostasis by elimination of endogenously generated hazardous materials [4].

1.1.2 Functions of the complement system

The complement system is a highly integrated network of proteins located either in plasma or on cell surfaces; in 2010 Ricklin *et al* [4] estimated that 48 proteins form the "core" of the system [1, 4]. Complement performs three main physiological activities: integration of innate and adaptive immunity [5]; defending against infections [6, 7]; and disposing of immune complexes and waste from inflammatory and other injuries [8-10] (Table 1.1)

Table 1.1 The three main physiological activities of the complement system (adapted from [1]).

Activity	Complement Protein Responsible For Activity
Host defense against infection	
• Opsonization	• Covalently bound C3b (and iC3b) and C4b; receptors on erythrocytes and phagocytes.
• Chemotaxis and activation of Leukocytes	• Anaphylatoxins (C3a, C4a and C5a); anaphylatoxin receptors on various cells.
• Lysis of bacteria and cells	• Membrane attack complex (C5b, C6, C7, C8 and C9).
Interface between innate and adaptive immunity	
• Augmentation of antibody responses	• C3b (and iC3b) and C4b bound to immune complexes and to antigen; C3-fragment receptors on B- cells and antigen-presenting cells.
• Enhancement of immunologic memory	• C3b and C4b bound to immune complexes and to antigen; C3-fragment receptors on follicular dendritic cells.
Disposal of waste	
• Clearance of immune complexes	• C1q; covalently bound fragments of C3 and C4.
• Clearance of apoptotic cells	• C1q; covalently bound fragments of C3 and C4. • MBL and Ficolins, bind apoptotic cells.

The proteins that together make up the proteolytic cascade at the heart of the complement system were named C1 to C9, more or less according to when they were identified and in approximate accordance with the order in which they participate in the cascade. In addition, several proteases in the alternative pathway are called "factors", namely factors B, D and I (and factor H, which is not a protease but is a cofactor for factor I). Upon activation, several of the proteins are cleaved and the resulting fragments are then assigned a lowercase letter, e.g. C3a and C3b. The proteins may be organized into three interconnected activation pathways (see Figure 1.1 and "Activation of the complement system", Section 1.2).

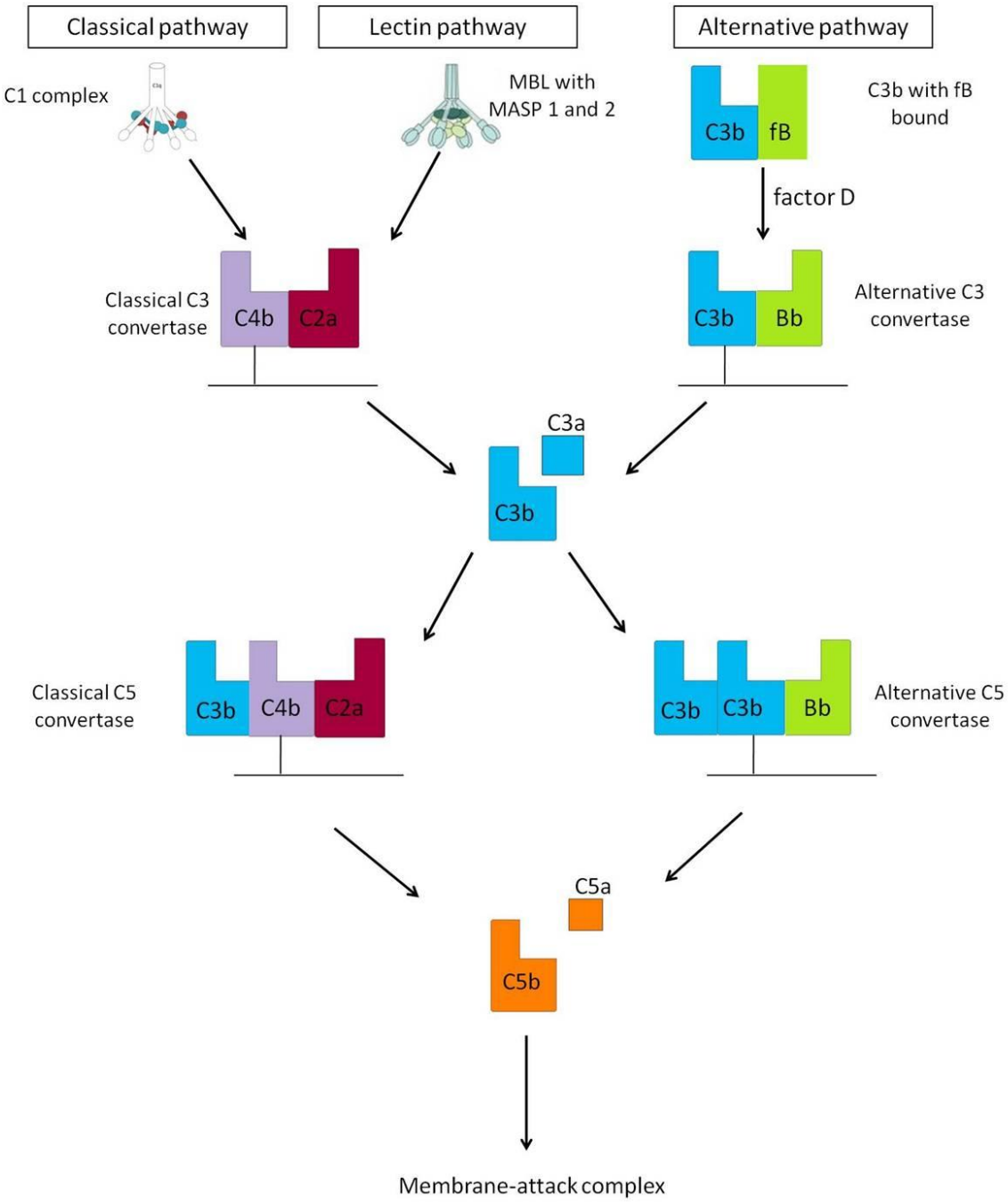


Figure 1.1 Three pathways of complement activation. The early stages of each pathway involve a series of proteolytic cleavages and macromolecular recognition events that allow an enzyme known as a "C3 convertase" to form. The event at which all three pathways merge is cleavage of C3 into C3a and C3b by the C3 convertases. Subsequently, C5 convertases assemble that catalyze activation of C5 generating C5a and C5b. The C5b fragment initiates self-assembly of terminal components to form the membrane attack complex (MAC). (Image redrawn form [1] and the MBL-MASP complex taken from Wallis group home page (http://www2.le.ac.uk/departments/biochemistry/staff/wallis/research) University of Leicester).

1.2 Activation pathways of the complement system

Complement activation is initiated by the classical, lectin or alternative pathways (Figure 1.1). All of the pathways involve initially inactive components that become proteolytically cleaved into their active forms. Each pathway is initiated by different molecules, but they converge at C3 cleavage and thereafter all can lead to production of anaphylatoxins (C3a and C5a), opsonins (C3b, iC3b, C4b) and, ultimately, the MAC.

1.2.1 The central role of complement components C3 and C3b

As mentioned above all three activation pathways converge at the cleavage of C3 to C3b. This is therefore the pivotal event in a complement-mediated response.

C3 (187 kDa) is part of the α2-macroglobulin family. It is composed of two chains - the α-chain (residues 650-1641, 110 kDa) and β-chain (residues 1-645, 75 kDa) - that are linked by a single disulphide bond and non covalent forces [11]. Electron microscopy [12] and crystal structures [13-15] (Figure 1.2) have revealed a large difference in the domain orientations between C3 and C3b. Of great significance is that the reactive thioester, which is buried in C3, becomes both exposed and activated in C3b allowing attachment to nucleophiles including hydroxyl groups on cell surfaces. The crystal structure of C3 (Figure 1.2 (A)) [13] revealed the presence of 13 domains. Eight of these are macroglobulin (MG) domains that form the core or body of the protein and have a key-ring arrangement; MG1-4 form one ring whilst MG5 and MG6 form a second half-ring which overlaps the first; MG7 and MG8 extend the second half-ring. The five remaining domains are the linker (LNK) domain located between MG1, MG4 and MG5, the anaphylatoxins (ANA) domain (residues 650-726), the CUB domain, the thioester domain (TED) that is inserted within the CUB, and finally the C345C domain.

C3b is generated from C3 by the proteolytic removal of the ANA domain from C3 which results in a truncated α-chain (residues 727-1641) whilst the β-chain (residues 1-645) is unaffected. The crystal structures of C3 and C3b (Figure 1.2) [14, 15] revealed for the first time, the details of the major rearrangement in domain orientations following excision of ANA. As mentioned above, the most important of

these rearrangements is the shift in position of the CUB and TED domain. Thus, in C3b the thioester (residues Cys988-Gln991 [16-18]) is exposed and hence poised to attach to the surface while in C3 these residues are buried at a position 85 Å away from where they are in C3b. The TED domain in C3b has also changed its conformation to be C3d-like [19] rather than C3-like [13] and the reactive residues of the thioester now form an acyl-imidazole intermediate that is highly reactive [20].

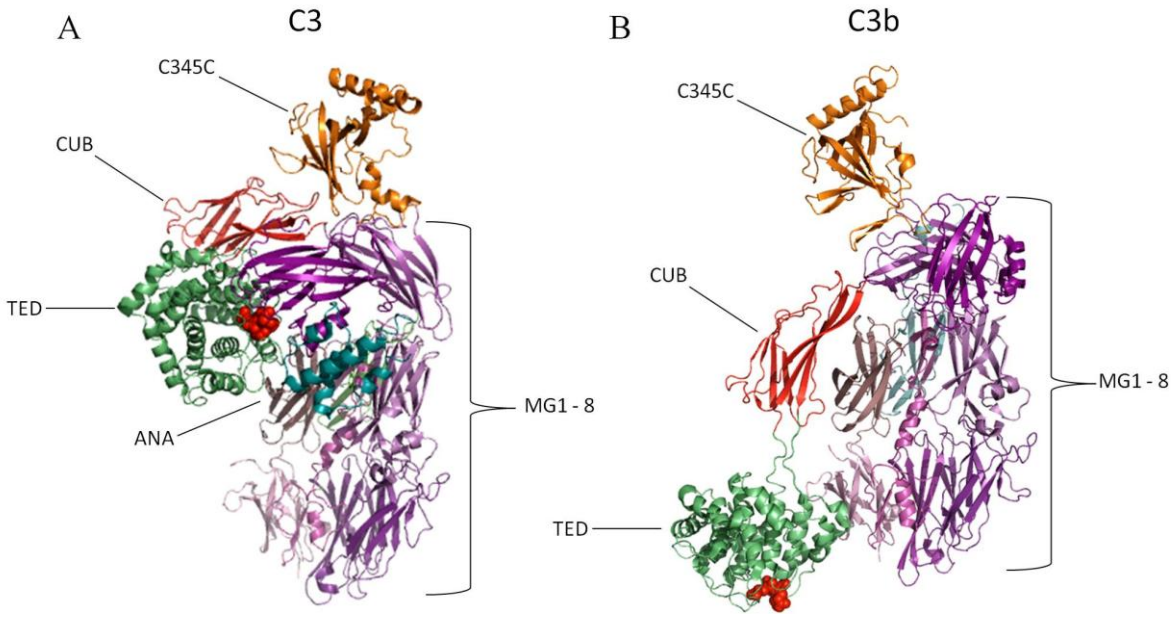


Figure 1.2 Crystal structures of C3 and C3b. (A) shows the crystal structure of C3 (PDB_ID 2A73, [13]) with the eight MG domains shown in different purple shades, ANA in cyan, CUB in red, TED in green with the thioester shown in red spheres, and C345C in orange. The crystal structure of C3b (PDB_ID 2I07, [14]) (B) shows the loss of the ANA domain and the large conformational change between the CUB and TED domains allowing the thioester to become solvent-exposed [13, 14].The structures were visualized using PyMOL (www.pymol.org).

1.2.2 Classical pathway

The classical pathway (CP) was, as the name suggests, the first pathway to be discovered. It is activated by C1; C1 is a molecular complex consisting of C1q associated with two molecules each of the serine protease proenzymes C1r and C1s. C1 is mainly activated through binding to the Fc portions of antigen-bound IgG and IgM antibodies [21-23]. Once bound to a molecular pattern on a target surface, a conformational change occurs to C1q, which is composed of 18 chains arranged in tightly associated trimers and resembles a "bunch of six tulips" with collagenous "stems" and globular "heads" [24] (Figure 1.3). Binding by multiple globular heads

of C1q to antibodies in an antibody-antigen complex produces a rearrangement of the stems. The resultant mechanical forces that are experienced by the C1r chains - which weave in and out of the collagenous stems of C1q - causes mutual cleavage and consequent activation by one C1r molecule of the other. Activated C1r then cleaves and activates the two C1s molecules. Activated C1s performs cleavage of C4 to C4b that (like C3b) can attach to surfaces via a covalent bond [24-26]. C4b is a platform for binding of C2, which in the context of the C4b:C2 complex is also a substrate for activated C1s. The resultant complex C4b.2a is the C3 convertase (C3-cleaving enzyme) of the classical pathway. Some of the C3b molecules generated in the subsequent cleavage event remain bound (or re-bind) to C4b.2a, forming the C5 convertase (C4b2a3b). This enzyme cleaves C5 into C5a and C5b; C5b binds to C6 to initiate the terminal pathway and formation of the membrane-attack complex (MAC).

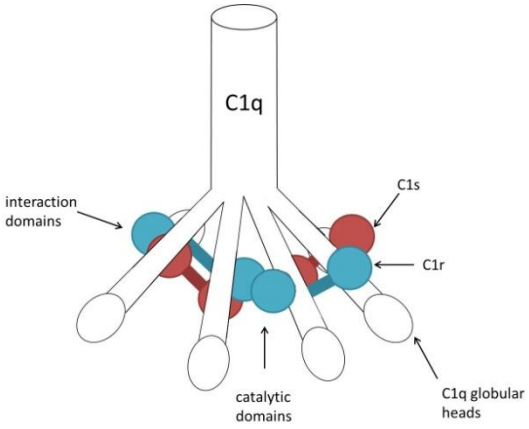


Figure 1.3 Structure of C1 complex. C1 contains one molecule of C1q and two molecules of C1r (blue circles) and C1s (red circles) which form the complex. C1r and C1s possess the interaction and catalytic domains. (Image redrawn from [27]).

1.2.3 Lectin pathway

The lectin pathway (LP) is similar to the CP. It also activates C2 and C4 but does so when mannose binding lectin (MBL) or ficolins binds to carbohydrates found on the surface of pathogens that include viruses, bacteria, yeast and parasites [4, 28, 29]. In serum both MBL and ficolins are in complex with MBL-associated serine proteases (MASPs) which are closely related to C1r and C1s of the CP [30-32]. MASP-2 is able to cleave both C2 and C4 and therefore promotes formation of C4bC2a, as in

the CP. On the other hand, MASP-1 cleaves C2 but not C4 [33, 34]. The function of MASP-3 has not yet been established.

1.2.4 Alternative pathway

The alternative pathway (AP) is triggered by the spontaneous low-level hydrolysis (also termed "tick-over") of the thioester bond buried in C3 (C3, C4 and C5 are all similar in structure and sequence but only C3 and C4 have thioester linkages) [35]. The hydrolyzed C3 forms C3(H_2O) (Figure 1.4, A), which undergoes a major conformational change (very similar to that of C3 when cleaved to C3b) and becomes a binding site for factor B (FB) [36]. Factor B in the context of the C3(H_2O)B complex is cleaved by factor D (FD) to generate the initiating C3 convertase (C3(H_2O)Bb). This (like C4b2a in the CP) cleaves C3 to form C3b and C3a [36].

As described in Section 1.2.1, cleavage of C3 by convertases into C3a and C3b exposes and activates a very reactive thioester group on C3b that allows it to covalently attach via an ester to a hydroxyl group on any nearby surface (in an analogous fashion to C4b that prefers amine nucleophiles). C3b (see Figure 1.2 (B)) is similar in structure to C3(H_2O) and is also a binding site for FB, forming the C3bB "proconvertase" complex. As with C3(H_2O)B, factor B in this complex is activated by FD to form the AP C3 convertase, C3bBb (Figure 1.4). This potentially surface-bound proteolytically active complex generates additional molecules of C3b that attach to the surface in the immediate vicinity and seed the formation of additional convertase complexes. In this way, a positive-feedback amplification loop is established resulting in rapid deposition of multiple molecules of C3b (up to 10^8 on one bacterial cell in ten minutes [37]). Some C3b molecules re-associate with C3bBb complexes resulting in formation of $(C3b)_2Bb$ and a switch of substrate to C5. Thus $(C3b)_2Bb$ is the C5 convertase of the AP (analogous to C4b2aC3b) while the resultant C5b is the starting point for assembly of the membrane-attack complex (see above). The C3b generated by the AP can also participate in forming the CP/lectin C5 convertase thus boosting all of the activation pathways. Importantly, the AP is very slightly activated (due to low-level C3 hydrolysis) at all times on every cell

exposed to serum, but amplification of C3b does not ensue due to the selective action of regulators. These are discussed in the next section.

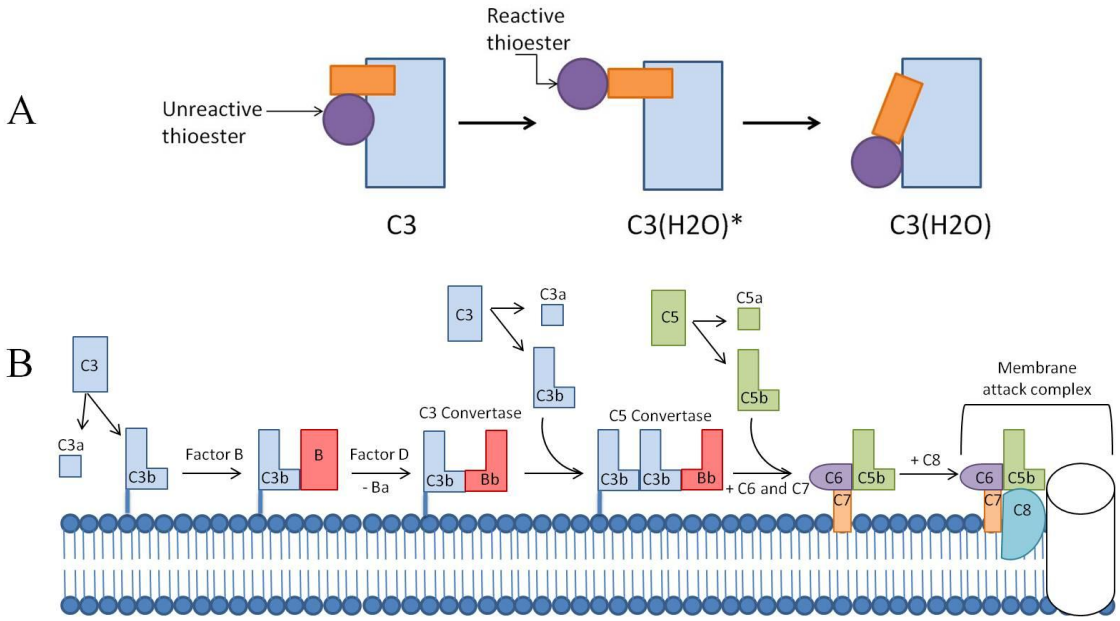


Figure 1.4 Activation of the AP. Shown in (A) is the spontaneous low level hydrolysis of C3 (via an initial reversible step that forms intermediate C3(H_2O)*) that results in the formation of C3(H_2O). The diagram in (B) shows the cleavage of C3 by C3 convertase and the attachment of C3b to the cell surface. The surface-tethered C3b then binds FB followed by cleavage by FD which results in the formation of a surface-attached C3 convertase. Some C3 convertases then bind a further C3b and form the C5 convertases, which nucleates the terminal events of the pathway resulting in the formation of the MAC. ((A) is re-drawn from [36] and (B) from [1]).

1.3 Regulators of complement

Complement regulation is vital to the wellbeing of the host organism. A balance between activation and modulation must be struck such that pathogens or damaged host cells are cleared, yet the level of activation must be controlled to avoid damaging tissue. In order to maintain this balance, a family of proteins called the regulators of complement activation (RCA) is present, both in the fluid phase and bound to cell membranes. These down-regulate the central steps in all three pathways that involve conversion of C3 to its activated form, C3b. Well characterized RCAs found in the fluid phase are FH and C4b-binding protein (C4BP) while membrane-bound regulators include complement receptor type 1 (CR1, also called CD35), membrane co-factor protein (MCP or CD46), decay acceleration factor (DAF also called CD55) (Table 1.2). Non-RCA members

including properdin, protectin (CD59) and C1-inhibitor are also important in complement regulation The fluid-phase RCAs have the important ability to act specifically to protect self surfaces (including ones not protected by a membrane) as opposed to bacterial and other foreign surfaces [38, 39].

The RCA proteins consist mainly of one type of domain; the complement control protein (CCP) module (also called a sushi domain or short consensus repeat). Factor H consists entirely of CCP modules. C4BP, MCP, and CR1 contain various numbers of CCPs along with: a linking chain (in C4BP that interlinks seven CCP-containing chains); transmembrane and intracellular domains (in CR1); a serine-threonine-proline-rich domain plus transmembrane and intracellular domains (in MCP) or a serine-threonine-proline-rich domain and GPI-anchor (in DAF) [38]. Table 1.2 summarizes the main proteins found in the RCA family.

Table 1.2 Summary of the regulators of complement activation.

Name	Abbreviation	Fluid Phase/ Surface bound	Number of CCPs	Ligands	Function
Factor H	FH	Both	20	C3b and C3d	Cofactor activity for C3b and accelerates decay of C3 convertase
Factor H-like protein 1	FHL1	Both	7	C3b	Cofactor activity for C3b and accelerates decay of C3 convertase
C4-Binding Protein	C4BP	Both	8	C4b	Cofactor activity for C4b and accelerates decay of C3 convertase
CR1	CD35	Membrane-integral	30	C3b, C4b, iC3b (and C1q)	Cofactor activity for C3b and C4b and accelerates decay of C3 and C5 convertases
CD46	MCP	Membrane-integral	4	C3b and C4b	Cofactor activity for C3b and C4b
CD55	DAF	GPI-anchored	4	C3 convertase	accelerates decay of C3 and C5 convertases

1.4 Factor H

Factor H (FH) was first identified in 1965 as β1H globulin by Nilsson and Mueller-Eberhard [40]. The gene encoding FH is *CFH* which is part of the RCA gene cluster on chromosome 1q32 [41]. FH itself is a single polypeptide chain composed of 20 CCP modules containing approximately 60 amino acid residues each; it has 1213 residues and eight N-glycans with a total mass of about 155,000. It is found in plasma at a concentration between 116-562 μg/ml [42, 43] but recently estimates have been lowered to 150-300 μg/ml [44].

1.4.1 Structural properties of Factor H

The 20 CCP modules (Figure 1.5) that form FH are attached to their neighbors via short linkers that seem to vary in flexibility. Since the N- and C-termini are at opposite poles of each of the spheroidal CCP module structures, modules within FH have been compared to "beads on a string". No crystals of FH have been reported in the literature while the available electron microcopy [45], analytical ultracentrifugation and small angle x-ray scattering (SAXS) study [46] point to a flexible overall structure that is, nonetheless, more compact (about half as long), on average, than would be the case if the protein were fully extended. Upon binding to C3b, FH is hypothesized to form a bent-back conformation that allows the N-terminus and C-terminus to interact with the same C3b molecule [47-49] but this is unproven.

All CCP modules that have been studied to date share a similar globular structure (40 Å by 15 Å by 10 Å) that is stabilized by disulphide bonds formed between four consensus cysteine residues in the pattern, CysI- CysIII, CysII- CysIV and a highly conserved set of hydrophobic residues – including a near-invariant Trp - that contribute to a buried core (Figure 1.5) [47].

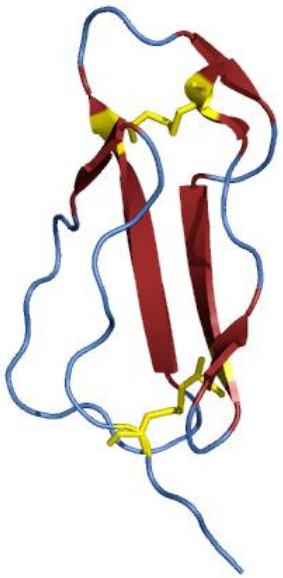


Figure 1.5 NMR derived structure of a CPP module. Cartoon diagram of CCP 16 of FH (PDB_ID 1HCC, [50]). β-strands are shown as red arrows connected by loops of non-standard secondary structure (blue). The four invariant cysteine residues are represented as sticks (yellow). The structure was visualized using PyMOL (www.pymol.org).

A multiple sequence alignment of all 20 FH CCPs (Figure 1.6) highlights highly conserved residues which include a proline residue at the third position after the first cysteine and tyrosines at a position two residues before (conserved in 14 out of 20 CCPs) and four residues (conserved in 12 out of 20 CCPs) after the second cysteine residue. Other common residues include Glycine and Glutamic acid.

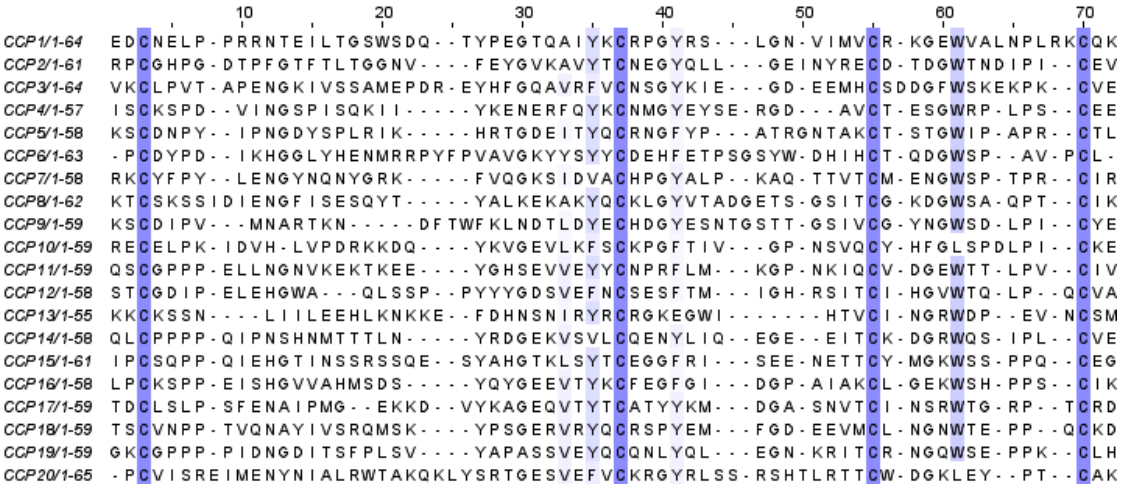


Figure 1.6 Multiple sequence alignment of the 20 CCPs of FH. Single-letter code used throughout. The invariant cysteines are highlighted by the dark blue box and the near-invariant tryptophan by the blue box. The tyrosine residues are also indicated (pale blue). The alignment was created using Jalview version 2 [51].

Currently there are 3D structures available for 17 out of the 20 CCPs (9, 14 and 17 are still unsolved) ,which have been determined by either X-ray crystallography or NMR spectroscopy using recombinantly expressed fragments containing between one and four CCP modules [50, 52, 53]. To date a full model for FH has yet to be determined but alongside the crystal and NMR models, techniques such as small angle X-ray scattering (SAXS), analytical ultracentrifugation and electron microscopy have helped to gain information regarding the overall shape of FH [54, 55].

The NMR-derived structures for CCPs 1-2 and 2-3, when over-laid on the common module, indicated that the N-terminal CCPs (1-3) adopt a highly extended "rod-like" shape [56]. A crystal structure of CCPs 1-4 in complex with C3b confirmed the "rod-like" shape of CCPs 1-3 but additionally revealed a kink between CCPs 3 and 4 giving CCPs 1-4 an L-like shape [57]. Data obtained using both SAXS [58] and FRET [59] confirmed the "L" shape of FH 1-4 . NMR-derived structures of CCPs 5 [60] and CCP 7 have been solved [61] whilst a crystal structure for CCPs 6-8 (in complex with sucrose octasulfate) showed an extended shape with a slight bend between CCPs 7-8 [62]. The central region of FH (CCPs 8-15) might act as a hinge or tether between the two termini of FH that are involved in binding to ligands (see Section 1.4.3) [63]. This central region contains CCPs that contain fewer residues (CCP 13 is the smallest at 51 residues) than average (60 residues) and these modules are joined by linkers that are longer in length when compared to the rest of FH (see figure 1.7).

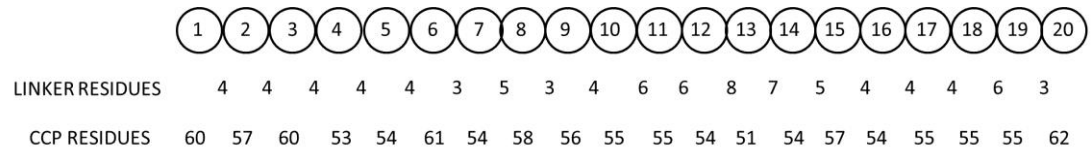


Figure 1.7 Linker lengths. The number of residues found in each linker is written below the CCP pairs. For each CCP the number of residues is written. Linker lengths and CCP residue data taken from (http://www.bionmr.chem.ed.ac.uk/bionmr/public_html/Residue_lengths.pdf).

Structures determined by NMR are available for CCPs 10-11, 11-12 and 12-13 while SAXS data is available for CCPs 8-15, CCPs 10-12 and CCPs 10-15 [52, 55]. The NMR-derived solution structures for CCPs 10-11 and 11-12 have intermodular

bends of 118° and 90°, respectively, while 12-13 is also a "V-shape", and features a bend of 80° (0° corresponds to a fully extended linear structure). Analysis of SAXS data using the ensemble optimization method (EOM) along with overlaid ensembles of solution structures for CCPs 10-12 suggests there is little flexibility between the modules. The lack of flexibility seen in CCPs 10-12 is also seen in the CCP pair 12-13 as relaxation data (T_1, T_2 and heteronuclear NOEs) recorded for the pair implied a relatively rigid conformation [55]. When the study was extended to include CCPs 10-13, a model based on the structure of the three pairs was created, showing that the modules formed an "out of plane zigzag structure" rather than the arc seen with CCPs 11-13. Analysis of the SAXS data for CCPs 10-15 with the model of CCPs 10-13 fitted as a rigid body yielded a confirmation that suggested CCPs 14 and 15 interacted very closely with CCP 13 resulting in a compact structure. The fitting of the NMR models to the SAXS data of CCPs 10-15 suggested that CCPs 13-14 form a near 180° bend allowing CCPs 14 and 15 to interact with CCP 13 and the 12-13 linker which gives a compact conformation to the central region [52, 55].

With regard to the C-terminus, both X-ray diffraction and NMR spectroscopy-derived models are available for CCPs 19-20 and there are crystal structures for FH 18-20 and for FH 19-20 in complex with C3d. The crystal structure [64, 65] and the NMR-derived structure [66] of CCPs 19-20 showed that a "rod like" shape was adopted. When the crystal model was extended to CCPs 18-20, a distinct "J" shaped was formed with an approximately 122° tilt between CCPs 18 and 19 [48]. This was not borne out, however by SAXS. Figure 1.8 shows all the NMR and crystal models that have been solved so far for FH.

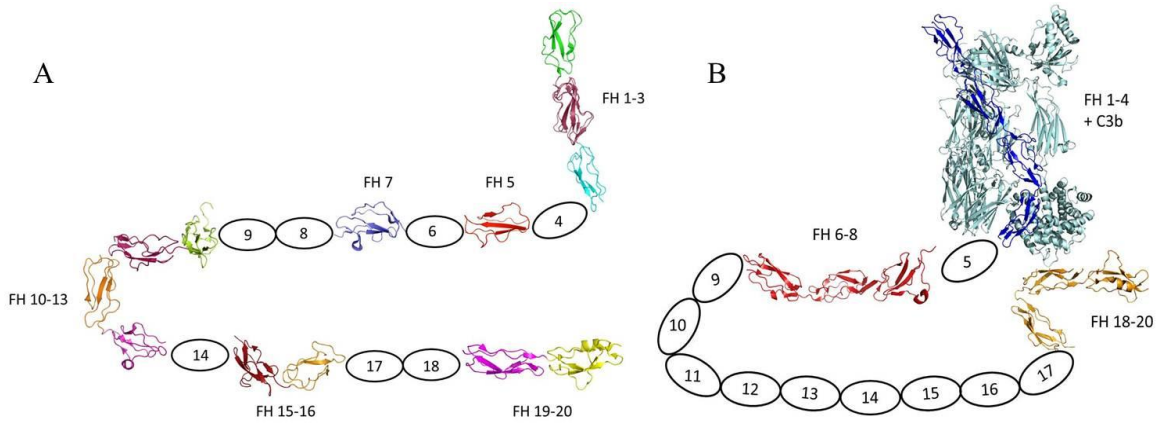


Figure 1.8 NMR and crystal-derived structural information for FH. (A) shows the CCPs that have had their structures solved using NMR whilst (B) show those modules solved by X-ray crystallography. The NMR section labeled FH 1-3 is composed of both NMR models of CCP 1-2 and CCP 2-3 aligned over CCP 2. PDB models used for NMR: CCPs 1-2 (2RLP, [56]), CCPs 2-3 (2RLQ, [56]), CCP 7 (2JGX, [61]), CCPs 10-11 (4B2R, [52]), CCPs 11-12 (4B2S, [52]), CCPs 12-13 (2KMS, [55]), CCPs 15-16 (1HFH) and CCPs 19-20 (2G7I, [64]).CCP 5 was taken from the CCP module database (http://www.bionmr.chem.ed.ac.uk/bionmr/public_html/ccp-db.html). For the crystal structure: CCPs 1-4 with C3b (2WII, [57]), CCPs 6-8 (2UWN, [62]) and CCPs 18-20 (3SW0, [48]).

1.4.2 Functional properties of Factor H

Factor H is the main soluble regulator of the AP and it protects host cells from damage that could be caused by activation of the complement system [40, 43]. Regulation by FH occurs in three ways. It acts as a cofactor for cleavage (of the alpha chain) by the serine protease FI of C3b into an "inactive" form, iC3b (Figure 1.9) with release of the small fragment C3f (this is known as cofactor activity). iC3b remains an opsonin but cannot bind factor B. Further cleavage by FI of iC3b α-chain (with other cofactors) generates C3c that is released and C3dg (that can be further degraded by other proteases to C3d), which remains bound to the cell surface by an ester linkage [36]. Factor H also blocks assembly of the AP C3 (and C5) convertases by competing with factor B for binding to C3b. Finally FH acts catalytically in accelerating the decay of any AP convertases that do manage to form, yielding C3b and Bb that cannot re-associate (this is called decay-acceleration activity) [36, 67]. Note that the convertases are already labile and prone to dissociate, with a half life of a few minutes, while FH reduces this to a few seconds. These regulatory functions of FH are contained within the four N-terminal CCPs such that a recombinant fragment consisting of FH CCPs 1-4 has both decay-accelerating activity and cofactor activity

in the fluid phase [68, 69]. Inactivation of C3b bound to a surface by FH (and FH/FI), however, is modulated by the chemical make-up of the surface to which it is bound. Discrimination by FH between self and non-self surfaces depends on CCP modules 5-20.

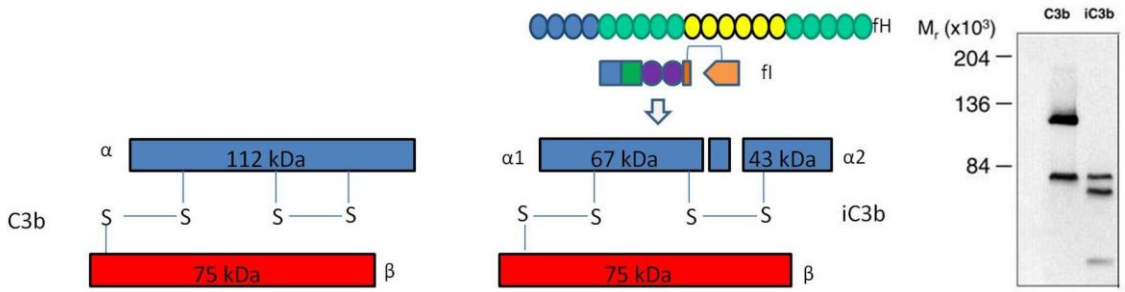


Figure 1.9 Cofactor assay. C3b contains two chains referred to as α-chain and β-chain linked by disulphide bonds. After incubation with FI and FH, the α-chain undergoes cleavage which results in the formation of iC3b. The α-chain cleavage can be visualized using SDS-PAGE followed by Coomassie staining or Western Blot; an example of a Western Blot is shown.

1.4.3 Binding sites of Factor H

Structure-function analysis of FH has highlighted binding sites for C3b and for polyanionic compounds such as glycosaminoglycans (GAGs).

C3b-binding sites

The main binding sites for C3b are located at both the N terminus and C terminus of FH [69]. At the N terminus, CCPs 1-4 interact with C3b. The crystal structure of C3b in complex with FH 1-4 (Figure 1.10) [57] showed a large buried interface of 4,500 Å^2. The interface between C3b and FH features four contact points that span the entire length of FH 1-4. Firstly the lower half of CCP 1 and the linker between CCPs 1 and 2 interacts with the new (following convertase cleavage) α-chain N-terminal residues (α'NT) and the MG7 domain through salt bridges and hydrophobic interactions. A second binding site is located on CCP 2 and is formed by an area of hydrophobic residues surrounded by hydrophilic residues which interact with the MG6 and CUB domains. The third site in CCP 3 makes contact with both the α- and β-chains as well as bridging the MG2 and CUB domains. Finally, CCP 4 bridges the MG1 and TED domains.

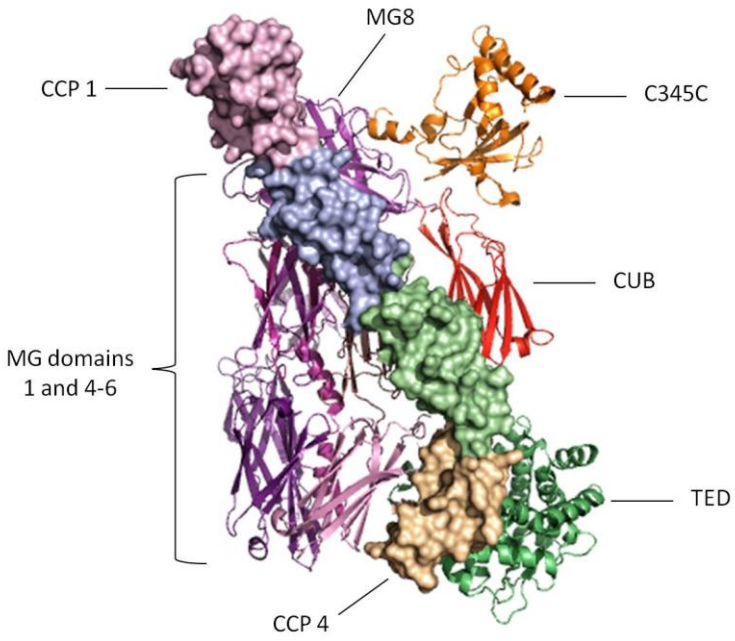


Figure 1.10 Structure of FH 1-4 in complex with C3b. The structure (PDB_ID 2WII, [57]) shows FH 1-4 represented as a surface (CCP 1 is pink, CCP 2 is pale blue, CCP 3 is pale green and CCP 4 is wheat) and C3b presented as a ribbon. The structure was visualized using PyMOL (www.pymol.org).

Binding to C3b at the C-terminus occurs through CCPs 19 and 20. As previously mentioned (section 1.2.1) the TED domain of C3b is "C3d-like" so the crystal structures of C3d in complex with FH 19-20 (Figure 1.11) [65, 70] can give useful information as to how the CCPs may interact with the TED domain of C3b. The complexes suggest that the main binding site for C3b is located in CCP 19 and the linker between CCPs 19 and 20 leaving CCP 20 free to interact with polyanions on the cell surface (see below) or, as suggested by Kajander *et al.* [70], bind a nearby C3d molecule on the cell surface. Interestingly recombinant FH 19-20 has been shown to discriminate between host and non-host cells and can inhibit full-length FH from binding to cell surface-bound C3b as well as inhibiting decay acceleration and cofactor activities on the cell surface. In fluid phase FH 19-20 has no effect on decay acceleration and cofactor activities [71].

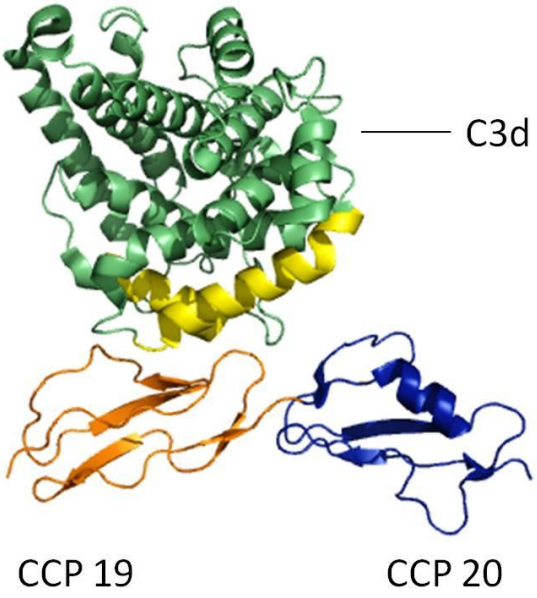


Figure 1.11 Structure of FH 19-20 in complex with C3d. The complex (PDB _ID 3OXU, [65]) highlights that most of the interactions occur between CCP 19 (orange) and the yellow section of C3d (teal). CCP 20 (blue) is free to interact with polyanions found on the cell surface. The structure was visualized using PyMOL (www.pymol.org).

A weaker binding site has also been suggested in the region of CCPs 6-8 [63] or CCPs 6-10 [72].

Polyanion binding

In order for FH to distinguish host cells from non-host it must be able to recognize molecular patterns on the cell surface. These patterns are mainly polyanionic markers (molecules that contain negatively charged patches) like glycosaminoglycans (GAGs) and sialic acid [73]. GAGs are linear polysaccharides of varying molecular weights that are found on the surface of all animal cells. They are usually linked to protein side chains. Sialic acid is a derivative of neuraminic acid and structurally contains a backbone of nine carbons. Figure 1.12 shows the structures of heparin and sialic acid.

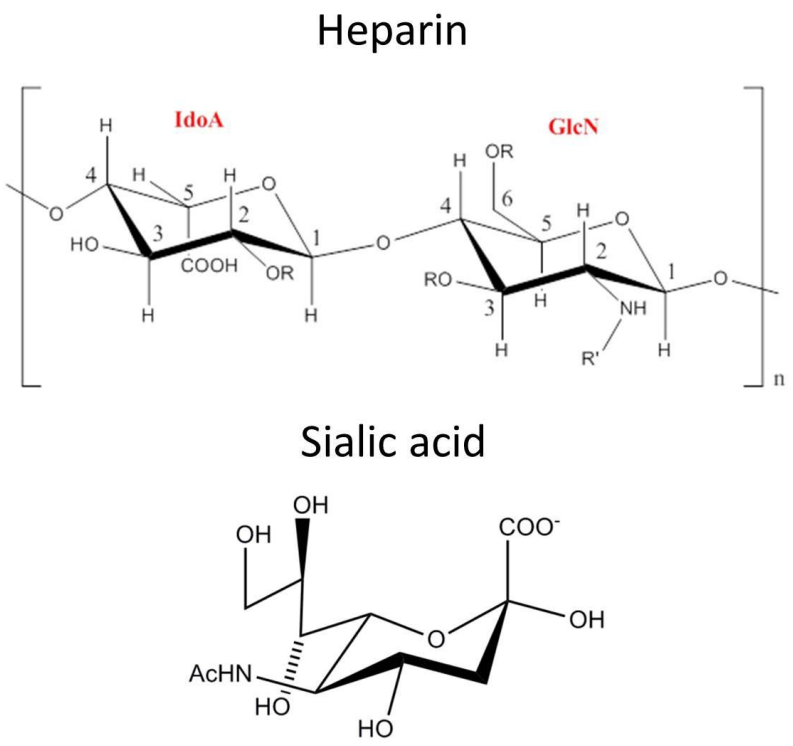


Figure 1.12 Structures of Heparin and Sialic acid. A typical heparin disaccharide unit (upper structure) comprising of IdoA and GlcN. R=$SO3^-$ or H, R'= $SO3^-$. or Acetyl group. For sialic acid (lower structure) the α-anomer is shown as this is the form sialic acid takes when bound to glycans. Structures were created using ChemDraw Std 12.0.

Sialic acids were first reported [74] to control the extent to which FH regulated amplification of surface-bound C3b. Studies have shown that FH could possibly contain three binding sites for heparin (which is used as a model for highly sulfated GAGs); CCP 7, CCP 20 and in the region containing CCPs 9-15 [63, 64, 66, 71, 75-78]. The heparin binding sites located in CCPs 7 and 20 are now unanimously agreed upon but the region containing CCPs 9-15 is uncertain. A study by Pangburn *et al.* [75] suggested that a binding site was located in CCPs 12-15 as fragments containing CCPs 6-15 and 12-15 bound to heparin agarose. The study also suggested that CCP 13 was the primary binding site based on cross linking, and consistent with CCP 13 containing a large number of basic residues compared to CCPs 12, 14 and 15. Two studies have been carried out to investigate whether CCP 13 was a binding site for heparin. The results from Blackmore *et al.* [76] did not support CCP 13 as a binding site for heparin as Δ13 (FH with CCP 13 deleted) bound to heparin-agarose affinity columns to a similar level as full length FH. Subsequently, Schmidt *et al.* [63]

showed recombinant fragments containing CCPs 12-13, 13, 11-14 and 13-15 did not bind to heparin.

In a separate study, another potential binding site was suggested for CCP 9. In 2006 Ormsby *et al.* [78], based on results with various recombinant fragments corresponding to the CCPs 8-15 region, suggested that either CCP 8 or 9 could potentially contain a binding site for heparin. That FH 8-9 and 9-11 both bound heparin suggesting that this site was in fact located in CCP 9. A later study by Schmidt *et al.* [63], however, found that recombinant FH 8-9 did not bind to a heparin-affinity column. These authors pointed out that the FH 8-9 construct expressed in the Ormsby *et al.* study had an N-terminal sequence artifact that contained two arginine residues and they showed that a peptide (EFTWPSRPSRIGTKT) that matched the non-native sequence of Ormsby's FH 8-9 (plus two native residues, K and T) had a weak affinity for heparin. In summary, no heparin binding site has been convincingly demonstrated to lie in the within the central region of FH corresponding to CCPs 8-15. Figure 1.13 summarizes the major binding sites for C3b and polyanions within FH.

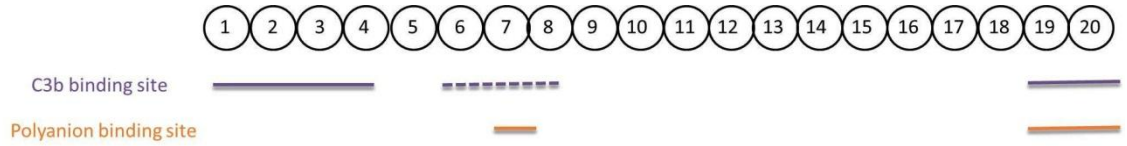


Figure 1.13 Major ligand-binding sites within FH. The N-terminal CCPs 1-4 and C-terminal CCPs 19-20 contain the major binding sites for C3b (solid purple line) whilst CCPs 6-8 contain a weak C3b-binding site (dashed purple line). The major polyanion-binding sites are located in CCPs 7 and 19-20 (solid orange line).

1.4.4 Factor H protein family

The FH protein family consists of seven multi-domain proteins which are solely composed of CCP modules; FH, Factor H-like protein 1 (FHL-1) and five factor H-related proteins (FHR-1, -2, -3, -4 and -5) [79]. FHL-1 is a splice variant of FH and is identical to the first seven CCPs of FH with an additional four amino acid extension at its C terminus (Ser-Phe-Thr-Leu). Due to the overlap with the first seven CCPs of FH, FHL-1 also shares cofactor and decay-accelerating activity as well as the ability to bind C3b and heparin [79, 80]. The FHR proteins (Figure 1.14)

are all encoded by their own genes which are located close to *CFH*. FHR-1 consists of five CCP modules with CCPs 1 and 2 being similar in sequence to CCPs 6 and 7 of FH and CCPs 3-5 being almost identical to the C-terminal CCPs 18-20 of FH. In human plasma two glycosylated forms of FHR-1 exist differing by one carbohydrate attachment site [81]. The smallest FHR is FHR-2, formed from four CCPs with CCPs 1 and 2 being similar to CCPs 6 and 7 of FH whilst CCPs 3 and 4 are similar to the C-terminus of FH. Like FHR-1 two glycosylated forms of FHR-2 exist - a single glycan and a 2-glycan form [82]. FHR-3 is composed of five CCP modules with CCPs 1-3 having high identity to CCPs 6-8 and CCPs 4 and 5 sharing similarity with CCPs 19 and 20 of FH [83]. The FHR-4B protein contains five CCPs with the first three being similar to FH CCPs 6, 8 and 9 while CCPs 4 and 5 of FH4B are similar (64% and 39% respectively) to 19 and 20 of the FH C-terminus [84]. The largest protein in the family is FHR-5 consisting of nine CCP modules with high identity to FH CCPs 6-7, 10-14 and 19-20 [85] (Figure 1.14). Functionally, all FHR proteins have been shown to bind C3b and polyanions but in general they lack detectable cofactor or decay-accelerating activity [79, 83, 86, 87]. The exception is FHR-5 which has been reported to display cofactor and decay-accelerating activity, although this is weak compared to FH [88]. Though they have neither cofactor nor decay-accelerating activity of their own, both FHR-3 and FHR-4B have been shown to enhance the ability of FH to perform these activities which was seen by the increase in cleavage of the α-chain of C3b into 68 and 41/43 kDa fragments visible on the SDS-PAGE gel [83].

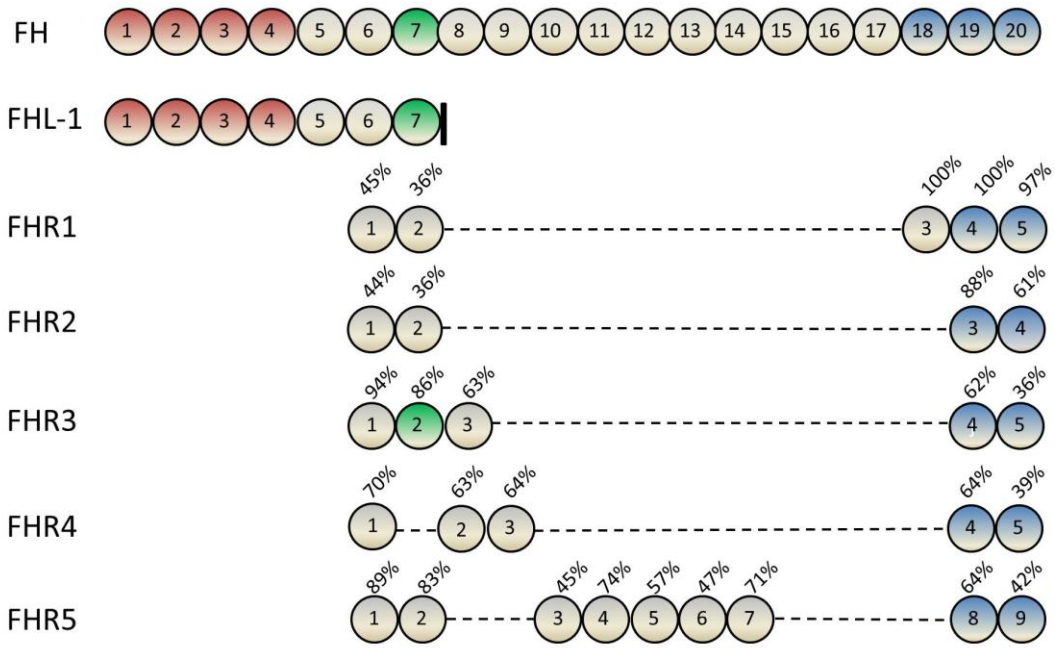


Figure 1.14 FH protein family members. The members of the FH protein family are composed entirely of CCP modules. The alignment is shown according to the highest percentage (%) of homology to FH. CCPs highlighted in red indicate cofactor and decay accelerating activity, green is for heparin binding and blue for C3b and polyanion binding. FHR4A is not shown. Image adapted from [79, 89, 90].

1.4.5 Factor H and disease

Polymorphisms, deletions and mutations found within the *CFH* gene cluster are associated with human diseases including the kidney diseases atypical haemolytic uraemic syndrome (aHUS) and membranoproliferative glomerulonephritis type II (MPGN II, also called dense deposit disease (DDD)), and the retinal disease, age-related macular degeneration (AMD) [43]. There are also links between FH sequence variations and susceptibility to bacterial infections [91]; moreover, since FH is one of the main regulators of complement, it is exploited by many pathogenic organisms as a means to avoid diminish complement attack [92].

Atypical hemolytic uremic syndrome

Atypical HUS is a rare but severe kidney disease that is characterized by thrombotic microangiopathy, haemolytic anaemia and acute renal failure [90]. There are two forms of aHUS, familial and sporadic. The majority of mutations linked to aHUS are heterozygous mis-sense mutations that are mainly found in the C-terminal region

with CCP 20 containing the most mutations (see Figure 1.15). At the N-terminus the highest number of mutations are located within CCP 1, including R53H [93] and R78G, which was found in a patient with sporadic HUS [94]. The SNP I62V also found in CCP 1 is linked to aHUS as well as being protective for AMD and DDD [95].

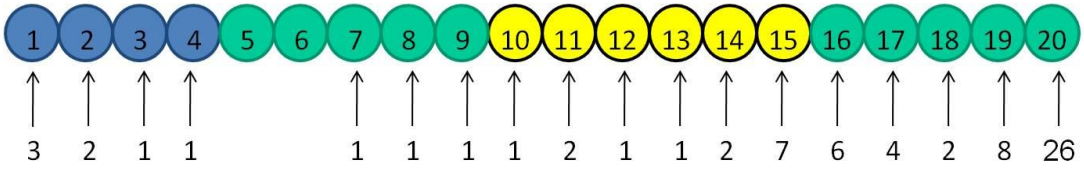


Figure 1.15 Number of aHUS mutations/SNPs located within the CCPs of FH.

In many cases, the mutant form of the protein is able to control complement activation in the fluid phase but unable to do so on self surfaces [65, 96, 97]. Nonetheless, in the familial form of aHUS, approximately 50% of individuals that carry an aHUS-linked mutation in FH will not developed the disease. Other triggers for developing aHUS may be required, such as an infection or pregnancy [98]. Alternatively, an additional genetic variant might contribute to the risk of developing the disease; this could be a mutation in one of the other complement proteins and it is now recognized that approximately 20% of patients will have mutations in more than one gene [99].

Membraneoproliferative glomerulonephritis type II

Another kidney disease associated with FH is MPGN II. The disease which is rare is commonly found in children and is characterized by the formation of complement-containing deposits in the glomerular basement membrane (GBM) that cause the kidney to become inflamed. In some cases deposits similar in structure and character are also found in the Bruch's membrane located below the retinal pigment epithelium (RPE) in the eye [100]. Unlike aHUS the mutations located in FH that are linked to MPGN II are not located in the C-terminal modules but are found throughout the protein. For example, a homozygous mutation located in CCP 4, ΔK224, was identified in siblings (sisters aged 10 and 5). The mutation is located on a surface exposed loop. It disrupts regulation of complement because the deletion of

this single amino acid causes C3b binding to be weaker, and both cofactor and decay acceleration activity are reduced. The deletion has no effect on binding to C3d or cell surface indicating the C-terminus of the protein is still fully functional [101].

Age-related macular degeneration

Age-related macular degeneration (AMD) is a degenerative disease of the macula which is one of the leading causes of vision loss, in the developed world, for individuals aged over 60. AMD can be classed in two forms; wet and dry. Wet AMD (also called neovascular AMD) results in loss of vision through the growth of fragile blood vessels that, if broken, allow blood and fluid to fill the void between the retina and RPE, resulting in damage to surrounding tissue. The dry form of AMD is characterized by atrophy of the RPE along with the development of deposits called drusen that form between the RPE and Bruch's membrane [102, 103] (Figure 1.16). Several studies have shown associations between AMD and the FH SNPs Y402H and I62V [95, 104-106]. The Y402 SNP located in CCP 7 can cause a six-fold increase in the risk of developing AMD in homozygous individuals whilst heterozygous individuals are 2.5 times more likely to develop the disease [107]. As CCP 7 contains one of the main binding sites for GAGs in FH the change from tyrosine to histidine may modify the ability of FH to bind to GAGs found on the cell surface [108]. In the case of I62V, which is found in CCP 1, the I62 variant is protective against AMD [95]. A quantitative study found that that when identical amounts of V62 FH or I62 FH were immobilized on a CM5 sensor chip within a surface-plasmon resonance instrument, and increasing concentrations of C3b was flowed across the chip, the I62 variant bound C3b more tightly (K_D = 1.04 μM ± 0.14 μM) than the V62 variant (K_D = 1.33 μM ± 0.12 μM) [95]. In fluid-phase cofactor assays, I62 FH was calculated to be 20% more active than V62 (using densitometry to quantify cleavage products) and in a surface-bound cofactor assay it was shown that an approximately twofold increase in V62 compared to I62 was needed to achieve 50% inactivation of C3b [95].

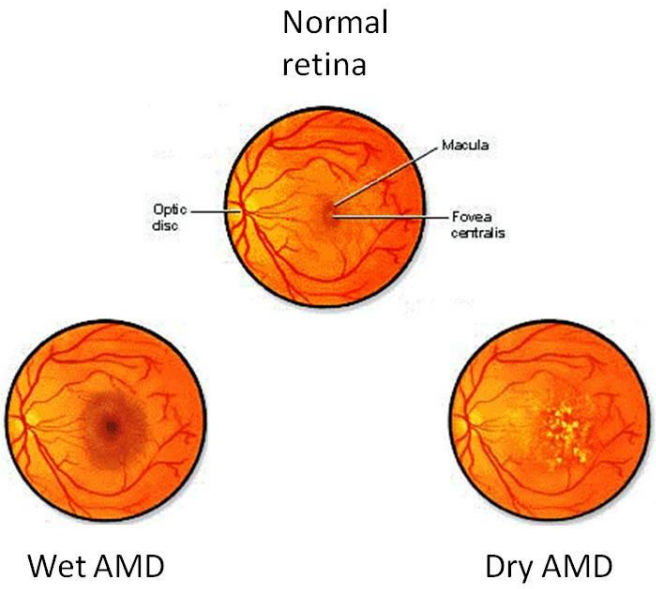


Figure 1.16 Age-related macular degeneration. Wet AMD is characterized by the formation of fragile blood vessels around the macular. Dry AMD is characterized by the formation of drusen deposits. Image adapted from http://www.beltina.org/health-dictionary/age-related-macular-degeneration-armd-symptoms-treatment.html

Capture of FH by pathogens

In order for pathogens to survive in the blood stream they must be able to avoid attacks from the immune system. The ability of pathogens to evade the complement system has been intensively studied over recent years [92, 109]. Three mechanisms of action have been identified: recruitment of complement regulators, inactivation of complement proteins caused by enzymatic degradation and inhibiting complement proteins by interacting with them directly [92]. A wide range of pathogens exploits FH by sequestering this regulatory protein on their surfaces. Examples of pathogens that capture FH are *Streptococcus pneumoniae* (Gram-positive), *Neisseria meningitides* (Gram-negative), *Candida albicans* (fungus) as well as *Echinococcus granulosus* (parasites) and *Human immunodeficiency virus* [92, 110].

Pneumococcus and pneumococcal surface protein C (PspC)

S. pneumoniae (Pneumococcus) is a bacterium found naturally in the upper respiratory tract and at low levels is considered a normal part of the upper respiratory tract flora (Javra). When the levels of bacteria are high it can lead to upper and lower

respiratory infections as well as causing serious infections like meningitis [111]. In order for pneumococcal diseases to progress several virulence factors are known to be involved, one of which is pneumococcal surface protein C (PspC, also referred to as choline-binding protein A (CbpA)) [112, 113].

Currently there are 11 different subtypes of PspC proteins and these can be divided into a further two groups based on how they anchor to the bacterial cell wall. Subtypes 1-6 are known as the classical PspC proteins and form binding group 1. The proteins employ a C-terminal choline binding domain that anchors them to the cell wall through a noncovalent interaction with the phosphorylcholine groups on the bacterial surface. Classical PspC proteins show similar structure and organization within the N-terminal regions, they all have a leader peptide and an N-terminal domain which is followed by one or two repeating domains then a proline-rich region (Figure 1.17). The second group is formed by subtypes 7-11 and uses an LPXTG motif which interacts covalently with peptidoglycans on the cell wall via a sortase-catalyzed transpeptidation reaction [114-116].

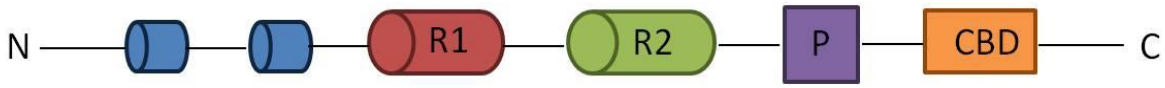


Figure 1.17 Illustartion of the domain organization of classical PspC proteins. The blue cylinders represent the N-terminal regions, R1 and R2 cylinders represent the repeating domains, the P square the proline-rich domain and the orange CBD rectangle the choline binding domanin.

PspC is able to bind FH which results in FH becoming attached to the pathogens surface, restraining complement activation and thus stopping complement mediated attack. Once bound to PspC, FH has been shown to remain functional. The interaction sites been the two proteins have been located to a 121 amino acid sequence found in N-terminal region of PspC and within CCPs 8-11 and 19-20 of FH [117, 118].

1.5 Investigating the structure of Factor H

In order to fully understand how FH carries out its functional roles we need to investigate it at a structural level. The main challenges associated with studying the structure of FH are its large size – 1215 residues – extended nature and flexibility.

High-resolution NMR and X-ray crystallographic methods have been used to obtained detailed structural models of several short segments of FH. Unfortunately both techniques have their drawbacks; NMR – even using high fields and taking advantage of isotopic labeling - is size-limited to three or four CCP modules, while crystallography requires well-ordered three-dimensional arrays (crystals) of the protein, which are less likely to form in the case of extended and flexible molecules. An additional technique that can be used alongside NMR and crystals is electron paramagnetic resonance (EPR) spectroscopy. This can provide useful information on structure and dynamics [119] for larger proteins and complexes that would be difficult or impossible to obtain using the more established high-resolution techniques.

EPR is particularly insightful when used in combination with site-directed spin labeling (more details in Section 1.5.3). This allows short and long-range distance measurements to be inferred. These distances can then be used as restraints when generating 3D models that have been built up from other information – such as models for the structures of FH and the C3:FH complex that are based on structures of fragments. The theory of EPR is similar to that of NMR. The major difference is that EPR makes use of the magnetic moment from unpaired electrons rather than nuclei [120].

1.5.1 Basic principles of EPR

Classically, electromagnetic radiation is thought of as magnetic (B_1) and electric (E_1) fields that are perpendicular to the direction of propagation. Electromagnetic radiation can also be represented by a stream of photons. The energy of a photon is given by $E= h\,v$, where h is Planck's constant (6.626×10^{-34} J s^{-1}) and v is frequency. When a photon has been either emitted or absorbed by an electron, atom or molecule, the energy of the total system must be conserved.

Every electron has a spin number ($s = {}^1/_2$) and a magnetic moment. Upon placing an unpaired electron in a magnetic field two spin states are created, anti-parallel ($m_s = +{}^1/_2$) or parallel ($m_s = -{}^1/_2$), and each alignment has a specific energy (due to the

Zeeman effect). The simplest energy level diagram for a particle containing a spin $^1/_2$ in a magnetic field is shown in Figure 1.18.

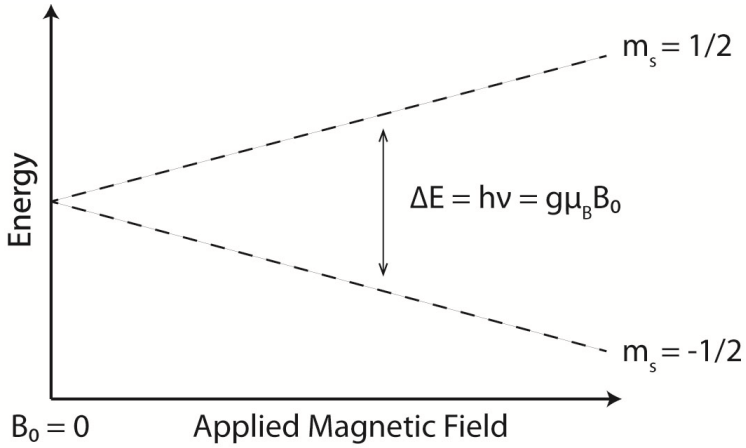


Figure 1.18 Energy level diagram. The energy levels are labeled $m_s = +^1/_2$ and $m_s = -^1/_2$ and by adjusting the magnetic field (B_0) the separation of the energy levels will change. When the frequency is adjusted so that the energy change $(\Delta E) = h\upsilon$, where υ is the frequency of the source, resonant absorption will occur. The second equation for ΔE in Figure 1.18 uses the electron g-factor (g), Bohr magneton (μ_B) and magnetic field B_0 [120].

1.5.2 Distance measurement theory

In all EPR methods, distance measurements depend on the magnetic dipole-dipole interaction between the magnetic moments $\vec{\mu}_A$ and $\vec{\mu}_B$ of the two spins, A and B. The energy of interaction (E) between the two magnetic moments is given by Equation 1

$$E = \frac{\vec{\mu}A \cdot \vec{\mu}B}{R^3} - \frac{3(\vec{\mu}A \cdot \vec{R}) \cdot (\vec{\mu}B \cdot \vec{R})}{R^5} \qquad (1)$$

Where R is the distance between A and B. The energy equation can be transformed to the quantum mechanical spin Hamiltonian using the relationship between the electron magnetic moment and the electron spin operator ($\vec{S}$), Equation 2

$$\vec{\mu} = -\gamma_e \hbar \vec{S} \qquad (2)$$

where the gyromagnetic ratio of the free electron is γ_e, and $\hbar$ is Planck constant divided by 2π. When the dipolar coupling is small in comparison to the Zeeman splitting of the electron spin states, the dipolar frequency can be calculated by Equation 3

$$\omega_{dip} = \frac{D_{dip}}{R^3}\left(1 - 3\cos^2\theta\right) \qquad (3)$$

with D_{dip} being the splitting constant. ω_{dip} can vary between $-2D_{dip}/R^3$ and $1D_{dip}/R^3$ for each rotation of the dipolar axis parallel ($\theta = 0°$) and perpendicular ($\theta = 90°$) with respect to the external magnetic field (B_O). The average value of the dipolar interaction is zero for molecules with random orientation that are rotating fast in comparison to the inverse coupling strength. In PELDOR and DEER experiments, the sample is generally frozen so the spectra are a superposition of randomly oriented bi-radicals and each consists of two lines split by dipolar coupling. The spectrum from a sample containing molecules orientated randomly is called a Pake pattern (Figure 1.19). Distance between the two edges of the Pake pattern relate to the dipolar splitting from molecules where $\theta = 0°$. The distance between the two peaks relates to the dipolar splitting for molecules where $\theta = 90°$ [121].

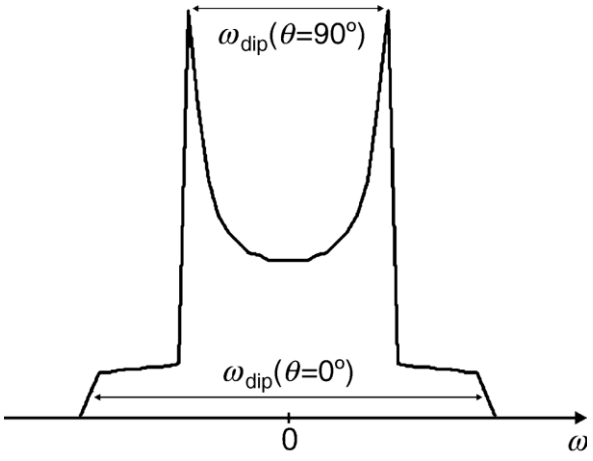


Figure 1.19 Pake pattern with peaks corresponding to 90° and edges to 0°. Image from [121].

1.5.3 Site-Directed Spin Labeling

Site-directed spin labeling (SDSL) combined with EPR has become a powerful tool for the investigation of protein structure and dynamics [122, 123]. A commonly used method for SDSL involves the introduction of a nitroxide side chain at a selected site which has been typically inserted using site directed-mutagenesis. Typically, cysteine-substitution mutagenesis is used (while any native cysteines are replaced by alanines or serines), followed by modification via a covalent interaction, of the sulfhydryl group with a specific nitroxide reagent, [122, 124] (more details in next section).

An alternative method is to incorporate an unnatural amino acid (e.g. *p*-acetylphenylalanine) into the protein sequence, using amber-codon suppression, which allows site-specific labeling [125]. In this approach, a stop codon is inserted into the DNA at the sequence position of interest. A "suppressor" tRNA which recognizes the stop codon and competes with release factor is then prepared and acylated with the unnatural amino acid; then together with the mutated DNA this is added to the appropriate translation system [126].

Additionally, enzymatic post-translational modification of proteins can be used to achieve site-specific labeling. The method works by inserting a short recognition sequence, typically at the N or C terminus of the protein, for a suitable enzyme. Sato *et al.* showed that the use of a transglutaminase enzyme and a short recognition sequence at the N-terminus (see Section 1.5.5) allowed a single fluorophore to be attached to a glutamine residue within the recognition sequence [127]. An advantage to this method is that it can be performed under mild conditions.

1.5.4 Modification of Cysteine

A thiol or sulfhydryl group is one of the strongest nucleophiles in a protein and is therefore available for selective modification with an assortment of spin-labeled reagents (Figure 1.20 using MTSSL as an example). Maleimide and iodoactamide derivatives that contain a nitroxide ring are suitable, but the most commonly used spin label is 1-oxyl-2,2,5,5–tetramethylpyrrol–3-yl-methyl methanethiosulfonate

(MTSSL). The modification of sulfhydryl groups in proteins is generally achieved at a pH between 7 and 9 and the thiolate ion is thought to be the reactive species [128].

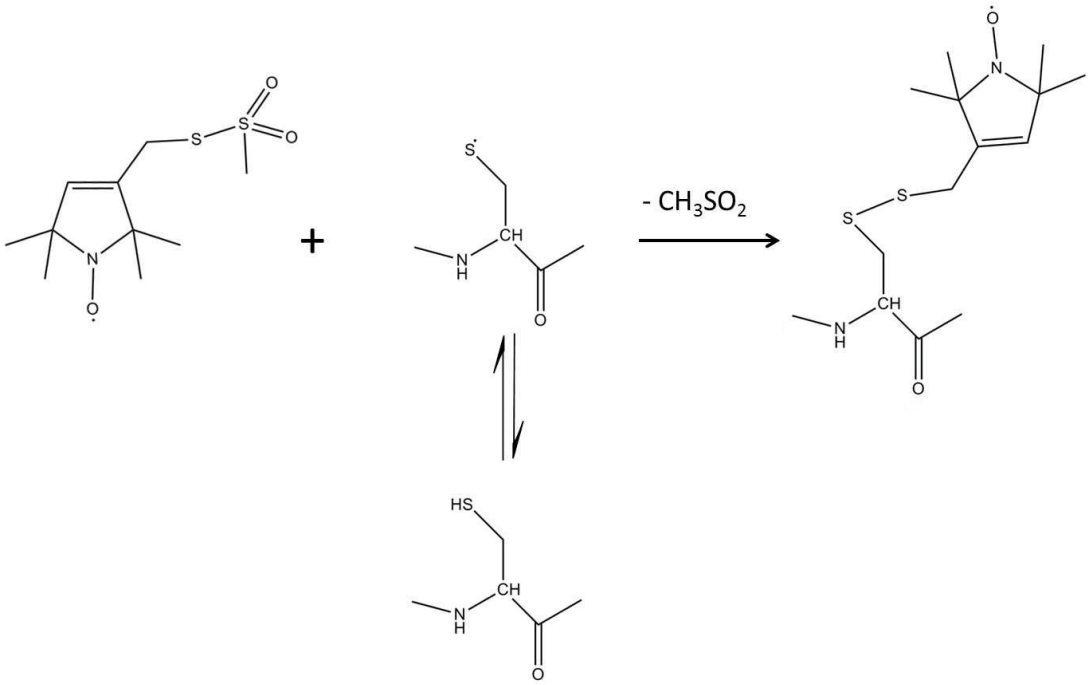


Figure 1.20 Reaction between cysteine and MTSSL. The cysteine is shown with the proton dissociated from the sulfhydryl group to form the thiolate ion, the reactive species.

1.5.5 Transglutaminase catalyzed labeling of proteins

One of the challenges of studying FH by EPR is that insertion of additional cysteines may disrupt the formation of the native disulphide bonds; enzymatic labeling circumvents this potential problem. Members of the transglutaminase (TGases) family catalyze the calcium (Ca^{2+})-dependent acyl transfer reaction between the γ-carboxyamide of a glutamine residue (acceptor) and a primary amine donor (Figure 1.21) [129].

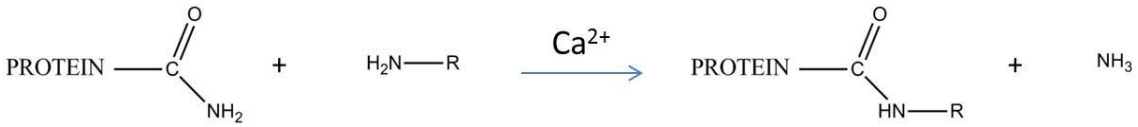


Figure 1.21 Transglutaminase reaction. The protein can be labeled at the glutamine residues when the R group is part of a primary amine in the calcium-dependent reaction.

The primary amine of lysine residues can act as a substrate, resulting in the generation of intra- or intermolecular cross links [130]. The actual requirements that allow a glutamine residue to be a substrate for TGase are not yet fully understood but TGases extracted from tissue display strict structural and sequence requirements [131]; to participate in acyl transfer, a glutamine residue needs to be accessible in a solvent-exposed position that is located in a flexible segment of polypeptide [132]. The specificity for the amine donor is less strict in that natural primary amines and non-biological substrates are recognized [131]. A study by Taki *et al.* compared the fluorescent labeling of glutathione S-transferase (GST) via chemical labeling techniques, with the TGase-catalyzed labeling approach. This study showed that TGase-catalyzed labeling was preferred as this procedure avoided the non-specific labeling of glutamine residues that occurred with the chemical labeling method. Moreover, TGase-labeled proteins were more likely to retain activity while the labeling reaction could, if necessary, be carried out at a temperature as low as 4° C, suitable for less thermally stable proteins like FH [133].

1.5.6 EPR in biological studies

EPR has been used to study structure and structural changes in proteins [134-138]. The development of pulsed EPR methods such as PELDOR (pulse electron-electron double resonance) and DEER (double electron-electron resonance) have allowed long-range distance measurements (up to 60 Å) to be made between two nitroxide labels or between a single nitroxide label and a paramagnetic centre. A collection of multiple distance measurements (based on multiple samples of individually site-specifically labeled proteins) allows triangulation and hence yields valuable structural information. For example it can shed light on the three-dimensional arrangements of proteins, subunits or protein domains within a large protein or complex. Such measurements also allow any structural changes upon binding with a ligand to be monitored [121, 139].

1.6 Aims of this work

Our working hypothesis is that the central region of FH, containing the smallest CCP (CCP 13) and the longest inter-modular linkers of seven and eight amino acid

residues, acts as a hinge which promotes a stable conformation in which the N and C termini engage simultaneously with C3b. Structural studies [52, 55] of the central segment of FH have suggested that CCPs 10-15 do indeed adopt a compact structure with little intermodular flexibility. Functionally, this region contains no known major binding sites for C3b or GAGs, consistent with a purely structural or architectural role [63]. Three theories have been advanced to explain the interaction of FH with C3b; (i) FH engages, 1:1, with a single C3b molecule located on the cell surface [65]; (ii) FH molecule is bound, 1:2, to two neighboring C3b molecules; (iii) FH can bind (1:1:1) to a C3b and a separate C3d molecule on the cell surface [70]. These theories are not mutually incompatible.

The first aim is concerned with the link between FH mutations and disease outlined earlier. As discussed, the N-terminus is necessary for both cofactor and decay accelerating activity. Two mutations that are linked to the kidney disease aHUS are located in CCP 1 and previous work has shown that these mutations reduce the ability of FH to bind to C3b or cause loss of cofactor or decay-accelerating activities. As most aHUS-linked mutations are heterozygous the work will aim to try and understand how the mutant protein disrupts the wild type protein thus resulting in the susceptibility to aHUS.

- Initial work will focus on the two mutations in CCP 1 (in the context of FH 1-4) and the development of a fluid phase cofactor assay which would allow both homozygous and heterozygous to be functionally compared.
- This will lay the groundwork for experiments in the context of full-length FH which will take into account the architecture of the intact protein.

As mentioned previously, full-length FH is too large and, probably, flexible for NMR or crystallography and hence a new approach is needed. SAXS has been successfully applied to FH as well as to fragments of FH but it does not allow domain arrangements to be reliably inferred in the absence of complementary data.

The second aim of the work was to lay the foundations for deploying EPR as a means of testing hypotheses regarding the “bent back” structure of FH and its engagement with C3 degradation products. In order to achieve this:

- Modified versions of FH 12-13 were produced in recombinant form at a high level to develop labeling strategies and record initial distance measurements.
- Apply this approach to longer fragments (such as FH 10-15), and to use the resultant distance measurements as possible restraints in NMR and SAXS-based models.

The overall goal of this strand of the project was to establish proof-of-principle by showing that EPR-derived distance measurements can be obtained for FH. Eventually – if all of the hurdles could be overcome - EPR could be immensely valuable in defining the overall architecture of full-length FH, free and bound to C3b

The third aim of the project also involved studies of FH 12-13. The uncommonly long linker (of eight amino acids) between these central CCPs forms a "mini-domain" between them that imposes a rigid V-shape. According to our hypothesis, this architectural feature at the centre of FH must be important for cooperation between the N- and C-terminal binding sites for C3b/GAGs. Thus the intention was to test this hypothesis by:

- Mutating the linker in such a way as to change the flexibility between domains, and to assess the functional and structural consequences.
- Quantify the effects of such mutations on flexibility (and module integrity) by working on the easy-to-study FH 12-13 module pair.
- Examine at a lower resolution (with SAXS) the structural effect of 12-13 linker mutations in the setting of FH 8-15 and FH 10-15 fragments.
- Study the functional outcome of these mutations by making them in the context of the full-length protein.

Chapter 2 Methods and Materials

For compositions of buffers and media, and for vector maps and sequencing primers, please see Appendices.

2.1 Amplification, cloning, screening and mutagenesis of DNA

2.1.1 Polymerase chain reaction (PCR).

PCR was used for the amplification of DNA sequences, and in the screening of colonies and sequencing of clones. Site-directed mutagenesis was also performed using PCR. All primers used in this study were either ordered from Sigma Genosys or Invitrogen and were shipped de-protected, and de-salted.

2.1.1.1 Amplification of codon-optimized factor H fragments.

A plasmid containing factor H-encoding DNA that had been codon optimized (by GeneArt) for expression in *Pichia pastoris*, was available in the group [140]. This served as the template DNA from which most of the fragments used in this study were generated. Two sets of oligonucleotide primers were designed. The first set allowed the amplification of the desired DNA sequence; the second set added the restriction enzyme sites (5' *Pst*I and 3' *Xba*I) as well as codons for any additional cysteine or tag sequence required for labelling experiments. Table 2.1 shows the sequences of all primers used in this study.

Table 2.1 Sequences of oligonucleotide primers.

Name	Sequence (5' – 3')	T_m (°C)	Restriction Site
CCP10 SNUG FOR	AGAGAGTGCGAGTTGCCAAAGATC	68.7	None
CCP12 SNUG FOR	ACTTGTGGTGACATTCCAGAATTGG	68.7	None

CCP13 SNUG REV	GGAACAGTTAACTTCAGGGTCCCA	67.9	None
CCP10 PST1 FOR	GATAAT*CTGCAG*GAAGAGAGTGCGAGTTGCCAAAGATC	79.1	*Pst*I
CCP10 N-CYS FOR	GATAAT*CTGCAG*GATGTAGAGAGTGCGAGTTGCCAAAGATC	79.4	*Pst*I
CCP10 N-TAG FOR	GATAAT*CTGCAG*GACCAAAGCCACAACAATTTATGAGAGAGTGCGAGTTGCCAAAGATC	86.2	*Pst*I
CCP12 PST1 FOR	GATAAT*CTGCAG*GAACTTGTGGTGACATTCCAGAATTGG	79.1	*Pst*I
CCP12 N-CYS FOR	GATAAT*CTGCAG*GATGTACTTGTGGTGACATTCCAGAATTGG	79.3	*Pst*I
CCP12 N-TTAG FOR	GATAAT*CTGCAG*GACCAAAGCCACAACAATTTATGACTTGTGGTGACATTCCAGAATTGG	86.2	*Pst*I
CCP13 PST1 FOR	GATAAT*CTGCAG*GAAAGAAGTGTAAATCATCCAACCTTATCATCTTGGAAG	79	*Pst*I
CCP13 XBA1 REV	AATAAT*TCTAGA*CTACTAGGAACAGTTAACTTCAGGGTCCCA	72.2	*Xba*I
CCP13 C-CYS REV	AATAAT*TCTAGA*CTACTAACAGGAACAGTTAACTTCAGGGTCCCA	73.9	*Xba*I
CCP13 C-TTAG REV	AATAAT*TCTAGA*CTACTACATAAATTGTTGTGGCTTTGGGGAACAGTTAACTTCAGGGTCCCA	81.6	*Xba*I
14 XBA REV	AATAAT*TCTAGA*CTACTACTCAACACACAAAGGGATGGACTGCC	75.6	*Xba*I
CCP 8 PST1 FOR	CATGAT*CTGCAG*GAAAGACTTGTTCCAAGTCCTCCATCGACA	83.7	*Pst*I

CCP 15 XBA1 REV	TAATAA*TCTAGA*CTACTATTCACATTGTGGTGGAGATGA CC	71.6	*Xba*I

Restriction enzyme sites are shown in italics. Codons for additional cysteine or tag sequences, which will be used in labelling experiments, are underlined. T_m is the melting temperature for each primer calculated using the DNA calculator from Sigma (http://www.sigma-genosys.com/calc/DNACalc.asp), FOR stands for forward and REV for reverse.

To generate the required coding sequences, a reaction was prepared containing 2 µl of template DNA (50-100 ng), 1 µl each of forward and reverse primers (25 µM stock), 1 µl deoxynucleoside triphosphates (dNTPs) (2.5 mM mix, Invitrogen), 1.5 µl dimethyl sulfoxide (DMSO), 5 µl of 10x Cloned Pfu buffer (Stratagene), 1 µl Pfu Turbo (DNA polymerase) (Stratagene) with the final volume being made up to 50 µl with double-distilled water. The thermal cycling parameters are shown in Table 2.2 and were implemented on an Eppendorf Mastercycler personal (Eppendorf, Germany) in 0.25 ml thin-wall PCR tubes.

Table 2.2 Amplification PCR program.

Step	Temperature (°C)	Time (min)	Cycles
Initialization	95	5	1
Denaturation	95	0.5	30
Annealing	T_m - 5°C	0.5	
Elongation	72	1	
Final elongation	72	10	1
Hold	4	-	-

The annealing temperature was set according to the lowest melting temperature (T_m) of the primer pair with 5 °C subtracted.

2.1.1.2 Screening of transformation colonies

Screening was performed on between eight and 20 colonies. The colonies were firstly re-streaked onto a freshly prepared low-salt LB agar plate containing Zeocin

(25 µg/ml final concentration) (Invitrogen) and left at 37 °C overnight. A small amount of the re-streaked colonies was then added to 20 µl "EB buffer" (10 mM Tris, pH 8.5) and heated to 95 °C for five minutes to cause the bacteria to lyse and to solubilize the plasmid DNA. After this, 2 µl of each lysate was transferred to a 0.25ml PCR tube containing 5 µl PCR master mix (Promega) and 2 µl of each primer. The primers used are shown in (Appendix B) and the thermal cycling parameters are summarized in Table 2.3.

Table 2.3 Program used for thermal cycling

Step	Temperature (°C)	Time (min)	Cycles
Initialization	95	1	1
Denaturation	95	0.5	39
Annealing	60	0.5	
Elongation	72	2	
Final elongation	72	4	1
Hold	4	-	-

2.1.1.3 Sequencing PCR.

Sequencing by PCR was performed using the same primers as used for screening by PCR (Appendix B). The reaction contained 250-300 ng of template DNA, 1 µl of either forward or reverse primer (10 µM stock) and 4 µl of ABI prism BigDye terminator v3.1 (Applied Bioscience), made up to 20 µl with double-distilled water. The thermal cycling program used for the sequencing PCR is shown in Table 2.4. The resulting PCR products were submitted to the Genepool automated sequencing service (University of Edinburgh) and the output was analyzed using BioEdit version 7.0.9 (http://www.mbio.ncsu.edu/bioedit/page2.html) or FinchTV version 1.4.0 (Geospiza).

Table 2.4 Sequencing PCR program

Step	Temperature (°C)	Time (min)	Cycles
Initialization	95	1	1
Denaturation	96	0.5	30
Annealing	50	0.5	
Elongation	60	1	
Hold	4	-	-

2.1.1.4 Site-directed mutagenesis

Clones expressing the desired DNA sequences (for FH 10-11, FH 11-14 and FH 12-13) were already available in the group. To perform *in vitro* site-directed mutagenesis the QuikChange II Site Directed Mutagenesis Kit (Stratagene) was used. Primers were designed according to the guidelines in the manual, which suggested they should be between 25 and 45 base pairs in length with 10-15 bases on either side of the mutation site, have a melting temperature greater than 78 °C and have more than 40% GC content with the ends being either a G or C. All primer sequences used are shown in Table 2.5. The PCR reaction contained 1 µl of template DNA (100 ng), 1.3 µl of forward and reverse primers (10 µM stock), 1.5 µl DMSO, 1 µl dNTPs (2.5 mM mix, Invitrogen), 5 µl of 10x Cloned Pfu buffer (Stratagene), 1 µl Pfu Turbo DNA polymerase and double-distilled water to make up to 50 µl. Table 2.6 shows the thermal cycling parameters used.

Table 2.5 Mutagenesis primer sequences.

Name	Sequence (5' – 3')	Tm (°C)	GC%
fH_10_15_CYS_5'	AGAGAGGCTGAAGCT*TGT*GGAGAATGCGAACTTCCT	79.8	50

fH_10_15_CYS_3’	AGGAAGTTCGCATTCTCC*ACA*AGCTTCAGCCTCTCT	79.8	50
11-14_CYS 5’	GAGAGGCTGAAGCT*TGT*CAATCATGTGGTCCACC	80.3	52.9
11-14_CYS 3**’**	GGTGGACCACATGATTG*ACA*AGCTTCAGCCTCTC	80.3	52.9
10-15_TTag_1+2 5’	TCTCCACCTCAGTGTGAA*CCAAAG*TAGTCTAGAAC	72.8	45.7
10-15_TTag_1+2 3’	GTTCTAGACTA*CTTTGG*TTCACACTGAGGTGGAGA	72.8	45.7
10-15_TTag_3+4 5’	CCTCAGTGTGAACCAAAG*CCACAA*TAGTCTAGAAC	74.1	45.7
10-15_TTag_3+4 3’	GTTCTAGACTA*TTGTGG*CTTTGGTTCACACTGAGG	74.1	45.7
10-15_TTag_5+6 5’	CAGTGTGAACCAAAGCCACAA*CAATTT*TAGTCTAGAAC	74.4	39.4
10-15_TTag_5+6 3’	GTTCTAGACTA*AAATTG*TTGTGGCTTTGGTTCACACTG	74.4	39.4
10-15_TTag_7 5’	GAACCAAAGCCACAACAATTT*ATG*TAGTCTAGAC	69.6	38.2
10-15_TTag_7 3’	GTCTAGACTA*CAT*AAATTGTTGTGGCTTTGGTTC	69.6	38.2
11-14_TTag_AA+1 5’	CAGTCAATACCACTCTGTGTT*GAACCA*TAGTCTAGAAC	72.3	42.1
11-14_TTag_AA+1 3’	GTTCTAGACTA*TGGTTC*AACACAGAGTGGTATTGACTG	72.3	42.1
11-14_TTag_2+3 5’	CAATACCACTCTGTGTTGAACCA*AAGCCA*TAGTCTAG	74.1	43.2
11-14_TTag_2+3 3’	CTAGACTA*TGGCTT*TGGTTCAACACAGAGTGGTATTG	74.1	43.2
11-14_TTag_4+5 5’	CTGTGTTGAACCAAAGCCA*CAACAA*TAGTCTAGAAC	73.7	41.6
11-14_TTag_4+5 3’	GTTCTAGACTA*TTGTTG*TGGCTTTGGTTCAACACAG	73.7	41.6
11-14_TTag_6+7 5’	CCAAAGCCACAACAA*TTTATG*TAGTCTAGAAC	68.2	37.5
11-14_TTag_6+7 3’	GTTCTAGACTA*CATAAA*TTGTTGTGGCTTTGG	68.2	37.5
12-13 3xGLY FOR	CAGTGTGTTGCT*GGT*GAC*GGTGGT*AAGAAGTGTAAATC	78.4	47.4
12-13 3xGLY REV	GATTTACACTTCTT*ACCACC*GTC*ACC*AGCAACACACTG	78.4	47.4

The bases highlighted in italics show the mutation part of the primer.

Table 2.6 Mutagenesis PCR program.

Step	Temperature (°C)	Time (min)	Cycles
Initialization	95	2	1
Denaturation	95	1	18
Annealing	55	1	
Elongation	68	10	
Hold	4	-	-

The QuikChange PCR product was then digested with 1.5 µl *Dpn*I (New England Biolabs – 20000 units/ml) for between two and four hours at 37 °C to allow the methylated template DNA to be digested before being transformed into chemically competent *E. coli* cells.

2.1.2 Restriction-enzyme digestion

All restriction enzymes were purchased from New England Biolabs and were used with the appropriate buffer and, if required, purified bovine serum albumin (BSA) that was supplied with the enzyme.

All the amplified DNA products and the pPICZαB vector were digested with *Pst*I and *Xba*I to generate overlapping ends. The digest mixture included amplified DNA or vector (50 µl), 1x New England Biolabs (NEB) buffer 3 (6 µl of 10x stock), *Pst*I and *Xba*I (1.5 µl of each 20,000 units/ml), 1x BSA (0.5 µl of 100x stock). NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$ and 1 mM DTT pH 7.9) was used as recommended by the NEB Double digest finder tool (http://www.neb.com/nebecomm/DoubleDigestCalculator.asp) The digests were carried out at 37 °C for three hours (amplified DNA) or overnight (pPICZαB). After this, the enzymes were heat-inactivated at 80 °C for 20 minutes followed by gel extraction.

In preparation for the transformation of the FH-encoding constructs into *P. pastoris*, the DNA was linearized using *Sac*I which cuts only in the pPICZαB vector (209 bp) and not the FH-coding insert. Each digest contained 440 µl of plasmid DNA, 6 µl *Sac*I (20,000 units/ml) 50 µl NEB buffer 1 (10x stock), 5 µl BSA (100x stock). NEB buffer 1 contains (10 mM Bis-Tris-Propane-HCl, 10 mM $MgCl_2$ and 1 mM DTT pH 7.0) Following incubation at 37 °C for six hours, the mixture was phenol-chloroform extracted to remove any remaining enzyme.

2.1.3 Dephosphorylation of digested vector

To prevent the re-ligation of the digested vector the 5' phosphate groups were removed. An aliquot of 50 µl of the digested vector was incubated with 1 µl (10 units) of calf intestinal alkaline phosphatase (CIP, New England Biolabs) and 5 µl of NEB buffer 3 for three hours at 37 °C, followed by gel extraction.

2.1.4 Ligations

All ligations were carried out using the T4 DNA Ligase (NEB) or the Rapid DNA Ligation Kit (Roche).

The following protocol was adopted when using the NEB kit that included the requisite buffer. Approximately 50 ng of digested pPICZαB vector was mixed with approximately 150 - 200 ng of the DNA insert and 2 µl of 10x "T4 DNA-ligase reaction buffer", made up to 19 µl with double-distilled water. After brief vortexing and centrifugation, the mixture was allowed to stand for ten minutes at room temperature. The T4 DNA ligase was then added (1 µl, 400 units/µl) and, following another brief vortexing and centrifugation, left to incubate overnight at room temperature. On the following day, 10 µl of the ligation mix was used to transform 50 µl of chemically competent XL1 Blue *E. coli* cells.

The following protocol was employed in the case of the Rapid DNA Ligation Kit. Insert DNA at a concentration of either 150 ng or 250 ng per µl was added to 50 ng of digested pPICZαB vector along with 1 µl of 5x "DNA-dilution buffer" and 5 µl of 2x "T4 DNA-ligation buffer" in a final volume of 10 µl. The reaction was then vortexed and allowed to stand for ten minutes at room temperature followed by the

addition of 1 µl T4 DNA ligase (5 units/µl). The reaction was then vortexed and left to incubate overnight at room temperature. On the next day, 4 µl of the ligation mix was used to transform 50 µl of chemically competent XL1 Blue *E. coli* cells.

2.1.5 Purification of plasmid DNA

To prepare plasmid DNA from bacterial cultures, a QIAprep Spin MiniPrep kit was used for small-scale amounts or a HiSpeed Plasmid Maxi kit for large-scale amounts (both from Qiagen). The QIAquick Gel extraction kit (Qiagen) was used to recover DNA from agarose gels and the QIAquick PCR Cleanup kit (Qiagen) was used to purify and concentrate any PCR products that were to be used as template DNA in other PCR reactions. All kits were used out according to the protocols supplied and the resulting DNA was stored at -20 °C.

2.1.6 Phenol-chloroform extractions.

To remove any remaining enzymes from the DNA prior to transformation into *P. pastoris*, phenol-chloroform extractions were used. The volume of DNA was measured and to this an equal volume of phenol:chloroform:isoamyl alcohol mixture (25:24:1, saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) was added. An emulsion was formed by inverting the tube several times, which was then centrifuged for 60 s in a bench-top micro-centrifuge (Eppendorf) at full speed (13,200 rpm). The resulting aqueous phase (top layer) was transferred to a fresh tube and the process was repeated. Finally, the aqueous layer was washed with an equal volume of chloroform, an emulsion was created and this was centrifuged as before. The DNA was recovered from the aqueous layer by ethanol precipitation.

2.1.7 Ethanol precipitations

Based on the volume of DNA recovered after phenol-chloroform extractions, 3 M sodium acetate solution, pH 5.2 (stock), was added to arrive at a final concentration of 0.3 M. Between 2 and 2.5 volume of cold ethanol was then added, followed by mixing, after which the sample was kept at -20 °C for at least an hour or overnight. The mixture was then centrifuged (Eppendorf centrifuge 5415R) for 30 minutes at 13,200 rpm, at 4 °C and the supernatant discarded. The pellet was washed with 400 µl 70% (v/v) cold ethanol and centrifuged for ten minutes at 13,200 rpm at 4 °C. The

supernatant was again discarded, and the pellet allowed to air dry to eliminate any remaining ethanol. Finally the pellet was re-suspended in 10 µl of doubled-distilled water and stored at -20 °C.

2.1.8 Quantification of DNA

The concentrations of DNA were estimated by spectrophotometry using a Biospectrometer (Eppendorf) to measure absorbance at a wavelength of 260 nm. The purity of the DNA was approximated by measuring the ratios of $A_{260\ nm/}\ A_{280\ nm}$ and $A_{260\ nm/}\ A_{320\ nm}$.

2.1.9.1 Preparation of chemically competent *E. coli* cells

Chemically competent XL-1 Blue cells were prepared according to the Qiagen method.

A starter culture was prepared containing 10 ml LB broth with 12.5 µg tetracycline and a small scrape of stock XL-1 Blue cells. These were cultured overnight at 37 °C, in a shaking incubator (200-250 rpm). The buffers were prepared as follows, and stored overnight at 4 °C. Buffer 1 contained 100 mM rubidium chloride, 50 mM manganese chloride:$4H_2O$, 30 mM potassium acetate, 10 mM calcium chloride and 15% glycerol, pH 5.8. Buffer 2 contained 10 mM MOPS, 10 mM rubidium chloride, 75 mM calcium chloride and 15% glycerol, pH adjusted to 6.8 with 1 M potassium hydroxide. After everything had dissolved both buffers were filter sterilized.

Next morning, 1 ml of the starter culture was added to 100 ml of pre-warmed LB broth containing 12.5 µg tetracycline in a 500ml baffled flask and cultured in the shaking incubator at 37 °C, 250 rpm, until the OD_{600} reached 0.5. The culture was then cooled on ice for five minutes and transferred to a centrifuge bottle. Cells were then collected by centrifugation at 4000 xg, 4 °C, for five minutes. The supernatant was discarded and cells re-suspended in 30 ml of cold buffer 1 and incubated on ice for 90 minutes. Cells were again collected by centrifugation (4000 xg, 4 °C, five minutes). The supernatant was again discarded and the cells re-suspended in 4 ml of buffer 2. Aliquots containing 50 or 200 µl were prepared in pre-chilled 1.5 ml tubes

which were then snap-frozen in liquid nitrogen and transferred quickly to the -80 °C freezer for storage.

2.1.9.2 Transformation into chemically competent *E. coli* cells

All transformations were carried out using 50 µl of XL-1 Blue super competent cells (Stratagene) or lab-made XL-1 Blue cells. The cells were mixed with either 1 µl of plasmid DNA, or 4-10 µl of ligation products and then incubated on ice for 30 minutes. Subsequently, cultures were heat-shocked by placing in a 42 °C water bath for 45 seconds then quickly transferred back to ice for two minutes. After heat shock, 250 µl of pre-warmed SOC medium (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 10 mM magnesium sulphate and 20 mM glucose per litre, Invitrogen) was added and the cells were incubated at 37 °C (250 rpm) for one hour to recover. The cells were plated onto low-salt LB agar containing 25 µg/ml Zeocin (Invitrogen) and incubated overnight at 37 °C.

2.1.10 DNA agarose gel electrophoresis

To prepare the gel, a mixture containing agarose and 1x Tris-acetate-EDTA (TAE) buffer was heated in a microwave oven until the agarose had completely dissolved. After addition of 5 µl SYBR Safe (Invitrogen) the mixture was poured into a tray containing a comb for the desired number of wells. Once the gel had set, a 100-bp or 1-Kb DNA size-ladder (New England Biolabs) was loaded alongside the DNA samples. The gel was run until the resolution of bands in the size-ladder was clear. Visualization of the gel was carried out using UV light on a transilluminator, and pictures were taken using a digital camera fitted with a UV filter.

2.2 Protein production in *Pichia pastoris*

2.2.1 Overview

Pichia pastoris was the eukaryotic expression host used in this work. It can perform the key post-translational modifications needed for production of factor H, namely disulphide bond formation. A mixed blessing is that it adds high-mannose glycans at

N-glycosylation sites [141, 142] that aid solubility but do not resemble the smaller and simpler sialylated glycans found on human factor H. *P. pastoris* is methylotrophic meaning it can exploit methanol as a sole source of both carbon and energy [142]. The first step in methanol metabolism is oxidation to formaldehyde and hydrogen peroxide catalyzed by the enzyme alcohol oxidase (AOX) in the peroxisome. The hydrogen peroxide is degraded to water and oxygen by catalase whilst a quantity of the formaldehyde leaves the peroxisome and is oxidised further to formate and carbon dioxide by two cytoplasmic dehydrogenases, with this reaction providing the energy for cell growth. Remaining formaldehyde is used to form cellular constituents by a cyclic pathway [142].

Two genes, AOX1 and AOX2, encode alcohol oxidase in *P. pastoris* with the two enzymes being 97% identical and sharing similar specificity [143, 144]. The majority of the alcohol oxidase activity is carried out by AOX1, which metabolizes methanol at a higher rate than AOX2. The strains used for protein production can be categorized based on which of the two alcohol oxidase genes has been deleted. The wild-type strain is Mut^+ (methanol-utilization plus phenotype). Strains lacking the AOX1 enzyme are termed Mut^S (methanol-utilization slow phenotype) and include the KM71H strain (Invitrogen) used in this work [141, 145].

P. pastoris can express foreign proteins intracellularly or can secrete them into the expression medium. Secretion of the foreign protein requires a secretion signal in order for the protein to be directed to the secretion pathway. The advantage of the foreign protein being secreted is that the first step of purification is easier as *P. pastoris* does not secrete high levels of native proteins [142].

2.2.2 Preparation of electrocompetent cells

A 5 ml starter culture containing a single colony of the *P. pastoris* strain KM71H and yeast-peptone-dextrose (YPD) medium was prepared in a 50 ml tube and placed in a shaking incubator for 24 hours at 30 °C, 200 rpm. Three 100 ml cultures containing fresh YPD medium were inoculated with 30-500 µl of this overnight culture, and were subsequently grown (overnight) and until the OD_{600} reached 1.3-1.5. The culture that had reached the required OD_{600} was then centrifuged at 1500

xg, 4 °C for five minutes and the resulting pellet was re-suspended in 100 ml ice-cold sterile water; this procedure was then repeated. Finally cells were centrifuged and re-suspended in 4 ml ice-cold 1 M sorbitol. A final centrifugation step was performed and the cells re-suspended in 200 µl of ice-cold 1 M sorbitol, then stored on ice until required. Cells were only used on the day that they had been prepared.

2.2.3 Transformation of *Pichia pastoris* by electroporation

Once the linearised DNA had thawed, 80 µl of the electrocompetent KM71H cells were added with agitation, and the mixture was transferred to a 0.2-cm ice-cold electroporation cuvette (Gene Pulser I, BioRad) and incubated on ice for five minutes. After incubation the cells were pulsed for 6 s at 1.5 kV and immediately afterwards, 1 ml of ice-cold 1 M sorbitol or yeast-peptone-dextrose-sorbitol (YPDS) broth was added to the cuvette. The contents were then transferred to a 15 ml tube and incubated at 30 °C, without shaking, for 2-3 hours. In the meantime, YPDS plates were poured containing either 100 µg/ml, 200 µg/ml or 300 µg/ml Zeocin (Invitrogen). After incubation, aliquots from the tube of 50 µl and 250 µl were spread out on the 100-µg/ml Zeocin YPDS agar plates, while the remaining volume was split equally between the 200-µg/ml and 300-µg/ml plates. Finally, the plates were incubated (30 °C for three to five days) until colonies were visible.

2.2.4 Small-scale flask expression tests

In order to see if the transformation into *P. pastoris* had resulted in detectable levels of protein production, small-scale trial expressions were carried out using the following protocol.

The transformed colonies were re-streaked onto a fresh YPDS agar plate, containing the same concentration of Zeocin as the original plate, then left for three days at 30 °C.

A 5 ml starter culture containing buffered glycerol-complex medium (BMGY) in a 50 ml tube was inoculated with a small amount of the re-streaked colony and incubated with shaking for 24-48 hours at 30 °C, 250 rpm. The 5 ml culture was then added to 100 ml of fresh BMGY medium in a 500 ml baffled flask and cultured for a

further 24-48 hours at 30 °C, 250 rpm. Finally, the cells were harvested by centrifugation at 3000 xg, 4 °C, for ten minutes and re-suspended in 25 ml buffered methanol-complex (BMMY) medium. The suspension was returned to the 500 ml baffled flask and incubated, with shaking (250 rpm), at room temperature. At 24-hour intervals, methanol was added, to a final volume of 0.5-1% (v/v) to induce protein expression for a total of three days. On the fourth day, supernatant was recovered by spinning at 3000 xg, 30 minutes and the cell pellet discarded. The supernatant was then filtered using a 0.22-µm sterile filter and brought to a final concentration of 5 mM ethylenediaminetetraacetic acid (EDTA) and 0.5 mM phenylmethylsulfonyl fluoride (PMSF) by addition of stock solutions. Protein expression was then checked using sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

2.2.5 Large-scale flask expression

Those colonies that had shown the best levels of protein production in the small-scale expression trials were then used in larger-scale expression attempts.

A single colony was added to 10 ml BMGY medium in a 50 ml tube and incubated at 30 °C, with shaking (250 rpm) for 24-48 hours. The 10 ml culture was then added to 100 ml of fresh BMGY in a 500 ml baffled flask and returned to the incubator for another 24-48 hours before being added to 400 ml of BMGY medium in a 2 L baffled flask. After a further 24-48 hours at 30 °C, shaking at 200 rpm, the cells were harvested by spinning at 3000 xg, 4 °C, for 15 minutes; the supernatant was discarded and the cell pellet re-suspended in 125 ml BMMY medium to induce expression. The temperature of the incubator was reduced to 20 °C and the agitation to 200 rpm. After initial induction the cells were fed methanol to 0.5% (v/v) final concentration every 24 hours for two days and then on the third day were fed twice. Finally, the cells were pelleted at 4000 xg, 4 °C for 30 minutes, and the supernatant filtered using a 0.22-µm sterile filter. As quickly as possible after harvest, EDTA (to 5 mM) (pH 8.2) and PMSF (to 0.5 mM) were added to reduce protease activity. The supernatant was stored either at 4 °C for immediate purification or -80 °C for long-term storage before purification.

2.2.6 Fermentation for ^{15}N enrichment

A fermentor was used when a recombinant protein were required be to be expressed in ^{15}N-enriched media for NMR studies. Fermentation, conducted on a Bioflow 3000 (New Brunswick Scientific), allowed the temperature, pH, agitation and air flow to be monitored and adjusted as required to increase the yield. All fermentations were performed in one-litre cylindrical fermentation vessels, using the following protocol.

A YPDS agar plate containing 100 µg/ml Zeocin was streaked with the desired *P. pastoris* clone and left at 30 °C for between three and five days until colonies were about 3 mm in diameter. Two 10 ml starter cultures were prepared by inoculating BMG medium with a single colony, and incubated at 30 °C for two days. The 10 ml cultures were then added to 130 ml BMG and incubated at 30 °C for a further two days.

While the cultures were growing, the fermentor vessel was prepared. The media (see Table 2.7) were prepared and poured into the vessel after all the probes were attached, then the entire unit was autoclaved.

Table 2.7 Basal salts and initial medium for 1-L fermentation

$CaSO_4$	0.75 g
$MgSO_4.7H_2O$	12 g
K_2SO_4	8 g
Glycerol	20 ml
1 M Potassium phosphate pH 6.0	60 ml
Double-distilled H_2O	Make up to 600 ml

After autoclaving all the probes were connected to the fermentor and the dissolved oxygen (DO) probe was charged overnight by setting the agitation to 200 rpm (and temperature to 30 °C) and the air flow to "on" so that the medium became saturated

with oxygen. Next morning the DO probe was calibrated, the base feed-line supplying 2 M potassium hydroxide was attached and the agitation/DO cascade set up. To make the media ^{15}N-rich for labeling, 7 g ^{15}N-ammonium sulfate (dissolved in 20 ml double-distilled H_2O) and 3.5 ml PTM1 salts (Amresco) were filtered (0.2 µm) into the media and 0.5 ml of autoclaved Antifoam 204 (Sigma Aldrich) was also added. The cells were then harvested and re-suspended in approximately 20 ml of 20 mM potassium phosphate, pH 6.0. To increase the density of the cells, they were grown in the presence of glycerol as a carbon source for one to two days prior to the addition of methanol. Before induction with methanol, the temperature was reduced to 15 °C and (for labeling) 1 g of ^{15}N-ammonium sulfate dissolved in 2 or 3 ml of double-distilled H_2O was filtered into the culture. Four hours later, gene transcription was induced by addition of methanol to a final concentration of 0.5% (v/v), and PTM1 salts were added. Once all of this initial addition of methanol had been consumed, further feeds of methanol – to 1% (v/v) - were given for three to four days; the timing of these feeds was based on the agitation rate and the amount of dissolved oxygen so as to avoid overfeeding or poisoning the cells.

Centrifuging the culture at 4000 xg for 30 minutes removed the bulk of the cell mass from the supernatant that was then transferred to fresh bottles for a second centrifugation step of 30 minutes at 4000 xg. After this the supernatant was carefully decanted and filtered (0.2 µm). Finally, stock solutions of EDTA and PMSF, to final concentrations of 5 mM and 0.5 mM respectively, were added in preparation for protein purification.

2.2.7 Ten-Litre Fermentation

All large-scale (ten-litre) fermentations were performed by Dr. John White. A 1-L BMGY (2 x 500 ml) culture grown for 48 hours at 30 °C, 200 rpm was used to inoculate the fermentor. To induce gene transcription 0.5 % (v/v) methanol and PTM_1 salts were added. After all of the initial methanol had been consumed, further feeds of methanol – to 1.5% (v/v) - and 1% tryptone (w/v) were given for three days; the timing of these feeds was based on the agitation rate and the amount of dissolved oxygen as before.

Initial removal of the cell mass was performed by centrifuging the culture at 4000 rpm, 4 °C for 30 minutes. The supernatant was then transferred to clean bottles and re-centrifuged for 45 minutes at 7000 rpm, 4 °C. After this the supernatant was carefully decanted and stock solutions of EDTA and PMSF, were added, as above, in preparation for protein purification.

2.3. Protein Purification

All pre-packed columns or resins were purchased from GE Healthcare. An ÄKTA FPLC system (GE Healthcare) was used to perform all chromatography steps.

2.3.1 Ion-exchange chromatography

As an initial purification step, an XK 26/20 column packed with 30 ml of either SP Sepharose fast flow (FF) resin (cation-exchange) or Q Sepharose FF resin (anion-exchange) was used. Crude supernatant from either flask or fermentor expression was diluted with double-distilled water and the pH adjusted to allow binding prior to being loaded onto the column which had been equilibrated with 20 mM sodium acetate (for cation-exchange) or 20 mM Tris-HCl (for anion-exchange) at the same pH as the cell culture supernatant (typically between pH 4 and 5 for cation-exchange chromatography and pH 8 or 8.5 for anion-exchange chromatography). In order to choose the best pH, the calculated pI (http://web.expasy.org/protparam) of each protein and the pH-range of the resin were taken into account. The column was eluted in two steps by first washing with binding buffer (same buffer used to equilibrate the column) for about five column volumes, and second applying a gradient from 100% binding buffer to 100% elution buffer over three column volumes, at a flow rate of 3.5 ml/min. The elution buffer was the same as the binding buffer but contained 1 M sodium chloride.

A Tricorn Mono Q 4.6/100 (1.7 ml column volume) was used to perform a higher-resolution anion-exchange step. This column was eluted at a flow rate of 2 ml/min with a two-step gradient consisting of 0–50% elution buffer over 20 column volumes followed by 50-100% elution buffer over three column volumes. The binding buffer was 20 mM glycine, pH 9.75, or 20 mM sodium carbonate pH 9.0; the corresponding elution buffer also contained 1 M NaCl.

2.3.2 Affinity Chromatography

In order to purify recombinant proteins containing a hexa-histidine tag, Ni^{2+}- affinity chromatography was performed. The columns used were either 1 ml HisTrap columns (GE Healthcare) or an XK 16/20 column packed with 25 ml IMAC Sepharose 6 fast flow (FF) resin (GE Healthcare). All columns were charged with $NiSO_4$. The buffers used for Ni^{2+}- affinity chromatography are shown in Table 2.8.

Table 2.8 Ni^{2+}- affinity chromatography buffers.

Buffer	Composition
Binding	20 mM potassium phosphate, pH 7.0 0.5 M sodium chloride
Wash	20 mM potassium phosphate, pH 7.0 0.5 M sodium chloride 20 mM imidazole
Elution	20 mM potassium phosphate, pH 7.0 0.5 M sodium chloride 500 mM imidazole

The wash buffer was only used when manually eluting the 1 ml HisTrap columns.

The elution of the 25 ml 16/20 IMAC column was performed at a flow rate of 3 ml/min, using a three-step imidazole gradient as follows: step 1, 0-8% imidazole over four column volumes; step 2, 8% imidazole for six column volumes; step 3, 8-100% imidazole in four column volumes. When HisTrap columns were used, samples were eluted manually.

2.3.3 Size-exclusion chromatography

Size-exclusion chromatography was performed on either a HiLoad 16/60 Superdex 75 prep-grade column at a flow rate of 1 ml/min, a Superdex 75 HR 10/30 column at a flow rate of 0.5 ml/min, or a Superdex 200 HR 10/30 column at 0.5 ml/min. For the HiLoad 16/60 Superdex 75 prep-grade column and Superdex 75 HR 10/30

columns the buffer was PBS (pH 7.4). The buffer used for the Superdex 200 HR 10/30 column was 20 mM potassium phosphate containing 500 mM NaCl, pH 7.4.

2.3.4 Concentration and buffer-exchange of protein samples.

After each step of purification the protein sample was concentrated using either a 6 ml or 20 ml Vivaspin (Sartorius or GE Healthcare) concentrator with a molecular weight cut off of 5,000, 10,000 or 30,000 Da as appropriate. The Vivaspin concentrators were also used to buffer-exchange the sample by spinning the sample to 0.5 ml then filling the concentrator with the new buffer, and repeating this process multiple times. Centrifugation steps were carried out at 3000 xg and 4 °C for varying amounts of time.

Mini-dialysis units (Thermo Scientific) (with a MW cut off = 3.5 kDa) furnished an alternative method for buffer-exchanging small volumes of protein sample. The protein was initially dialyzed into the required buffer for a period of a few hours before the buffer was renewed and the sample left to dialyze overnight. All steps were carried out at 4 °C with continuous stirring.

2.3.5 Estimation of protein concentration

The concentrations of protein were estimated using a Biospectrometer (Eppendorf) by measuring absorbance at 280 nm. The online Protparam tool from ExPASy (http://web.expasy.org/protparam) was used to calculate the extinction coefficient for each construct. The application of Beer-Lamberts law allowed the concentration of the sample to be calculated. Purity of the protein was approximated by the ratios of $A_{280\ nm}/A_{260\ nm}$ and $A_{280\ nm}/A_{320\ nm}$.

2.3.6 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

For SDS-PAGE, either 12% agarose or 4-12 % agarose gradient gels from the NuPAGE Bis-Tris range (Invitrogen) were prepared according to manufacturer's instructions. Samples were prepared by mixing the protein with 4x NuPAGE LDS sample buffer (106 mM Tris HCl, 141 mM Tris Base, 2% lithium dodecyl sulfate, 10% glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue G250, 0.175 mM phenol red, pH 8.5, Invitrogen) and, if reduced samples were required, 10x NuPAGE

sample-reducing agent (Invitrogen, 500 mM dithiothreitol) was added to a final concentration of 1x. Then the sample was heated at 95 – 100 °C for between two and five minutes. The sample was run out on the gel alongside the Precision-Plus protein ladder (10 – 250 kDa range) at 180 V until good separation of the ladder was achieved (approximately 50 minutes). Prior to staining, the gel was washed in water followed by heating, using a microwave oven, for 30 seconds; this was repeated three times to remove the SDS from the gel. Staining was performed using either Bio-safe Coomassie stain (Bio-Rad), which was left for one hour at room temperature, or lab-made Coomassie stain which was heated again in the microwave for ten seconds. To remove any excess stain the gel was washed in water for a few hours, with the water changed at least three times before being scanned.

2.3.7 Western blot

A western blot was used to confirm expression of full-length factor H. Firstly, the proteins were resolved as described in section 2.3.6 but the staining step was omitted and the gel was instead soaked in 1x transfer buffer (10x stock, Invitrogen). The gel was then transferred to a nitrocellulose membrane (BioRad) using an XCell II™ .Blot Module (Invitrogen) at 30 V for 1 hour in 1x transfer buffer. Pre-stained molecular weight markers allowed for the transfer to be monitored. After transfer the membrane was transferred to a 50 ml tube and blocked using 40 ml of blocking buffer (1x PBS with 5% milk) for four hours at 4 °C on a roller (Stuart scientific). After blocking the membrane was treated as follows:

1) Washed with 10 ml PBS for 10 minutes.
2) Given four washes (10 minutes each) with 10 ml PBS containing 0.05% TWEEN 20.
3) Given a final wash with 10 ml of PBS for 10 minutes.

The membrane was then incubated overnight at 4 °C on a roller with 10 ml of the primary antibody - *i.e.* Shp pAB to factor H, (Abcam) - in a 1:200 dilution in 5 ml of blocking buffer. After incubation the membrane was again rinsed as described above, the membrane was then incubated with the secondary antibody (Rb pAB to Shp IgG (HRP)) at a 1:20000 dilution in 5 ml of blocking buffer for 2 hours at 4°C

on a roller. The membrane was rinsed again as previously described. Development of the membrane was achieved using the SIGMA*FAST*™ 3,3′-diaminobenzidine tablets.

2.3.8 N-linked deglycosylation

Proteins that were suspected to contain N-linked glycosylation sites were deglycosylated either at the end of the test (small-scale) expressions or after the first purification step for large-scale expression. For test expressions, the protein sample was treated with 1 µl of a 1:3 dilution of Endo H_F (1,000,000 units/ml New England Biolabs) for one hour at 37 °C. The protein from the large-scale preparations (*i.e.* ~50 – 100 ml) was first adjusted to pH 5.5. This is the optimal pH for Endo H_F, and was only chosen providing the protein was stable at this pH. Alternatively, pH 6.0 was used, then the protein was incubated at 37 °C for six hours with 15 µl Endo H_F added along with protease inhibitors.

2.4 Protein Characterization

2.4.1 Mass spectrometry

2.4.1.1 Accurate mass

Samples were submitted to the Scottish Instrumentation and Resource Centre for Advanced Mass Spectrometry (SIRCAMS) at the University of Edinburgh for high-resolution liquid chromatography (LC) – MS, with the aim of acquiring an approximate mass. The samples were supplied in 2–5 µM concentrations in 20 µl of either 20 mM potassium phosphate, pH 6, or PBS. Data was collected by passage through an LC-MS U300HPLC 500 µM PSDVB monolyth column (Dionex) followed by Fourier-transform ion-cyclotron resonance (FT–ICR) mass spectrometry (Bruker Daltonics 12T SolariX Fourier Transform Ion Cyclotron Resonance Mass Spectrometer).

2.4.1.2 In – gel trypsin digest

Prior to performing the trypsin digest the protein samples were run on a 4 – 12% agarose gradient SDS-PAGE gel, which was subsequently washed thrice with 100 ml of H_2O. Staining was carried out for 30 minutes on a shaking platform (~ 50 rpm)

with InstantBlue protein stain (Expedeon) followed by three 10-minute washes to remove excess stain. The band to be analysed was then cut from the gel along with the smallest possible amount of polyacrylamide and placed in a 1.5 ml Eppendorf tube for same-day digestion, or storage (-20 °C).

Once the segment of gel had been excised it was washed with 300 µl of 200 mM ammonium bicarbonate (ABC) in 50% (v/v) acetonitrile (ACN) at room temperature for 30 minutes, and this was repeated three times to remove SDS and stain. The gel segment was then incubated in 20 mM dithiothreitol (DTT), 200 mM ABC in 50% ACN for one hour at room temperature to reduce the protein. Subsequently, the gel segment was washed three times with 300 µl of 200 mM ABC in 50% (v/v) ACN followed by incubation in 100 µl of 50 mM iodoacetamide, 200 mM ABC in 50% (v/v) ACN for 20 minutes at room temperature to alkylate the cysteines. The excised gel segment was then washed three times with 500 µl of 20 mM ABC in 50% ACN before being covered with 100% ACN. Once the gel segment had turned white, acetonitrile was removed and the gel segment allowed to dry (*e.g.* 10 minutes at 37 °C). The gel segment was then swollen in 29 µl of 50 mM ABC containing 2 µl of trypsin at 4 °C for up to one hour (or until swelling was evident); note that at 4 °C, auto-digestion of trypsin is avoided. Tubes were then sealed with Parafilm and incubated at 32 °C overnight (16 – 24 hours). The tube was then placed in a sonicating water bath for ten minutes to extract the digested protein. Finally, data was collected by SIRCAMS using electrospray ionization-mass spectrometry.

2.4.2 NMR

Protein samples for NMR analysis were concentrated and buffer-exchanged into 20 mM potassium phosphate, pH 6.6. The sample then had D_2O (Sigma Aldrich) added to a final volume of 10% of the sample volume to act as a frequency lock. Spectra (1D proton and 2D ^{1}H, ^{15}N HSQC [146]) were recorded on an AVANCE™ 18.8 Tesla (800 MHz) spectrometer (Bruker) fitted with a 5-mm triple-resonance TCI cyroprobe. The spectra were acquired by Juraj Bella, Elisavet Makou and Charalambos Panagos.

Relaxation methods

Spectra (T_1, T_2, and NOE) were recorded on an AVANCE™ 600 MHz spectrometer (Bruker) by Jurja Bella.

Processing and analysis of the data was performed by Dr Mateusz Maciejewski. Raw spectra were processed using the NMRPipe software [147]. All T_1, T_2, and NOE spectra were analysed using the CCPNmr Analysis software v. 2.2.2 [148]. The NMR assignments of the wild type protein (WT), as well as the short linker (SL), and the three-Gly-insert (3xGLY) mutants were carried over from the assignments associated with the PDB entry 2KMS. T_1 and T_2 values were extracted from the spectra by non-linear least square fitting the decaying intensities observed in the spectra, as implemented in CCPNmr Analysis [148]. In house C programs were used to calculate the diffusion tensors for the considered proteins, and to calculate their model free (MF) [149] and extended model free (EMF) [150] parameters. In these programs, the parameters were determined by conducting a coarse grid search in the parameter space of a given tensor model, followed by a Powell minimization [6] of the error function E. In all preliminary diffusion tensor determinations E was based on a sum of squares of differences (normalized to the experimental error) between the experimentally measured T_1/T_2 ratios and the T_1/T_2 ratios calculated for each residue n using the model being tested. In tensor refinements the differences in T_1, T_2, and NOE relaxation values were used in the same fashion to calculate E. Finally, the MF and EMF parameters were calculated for each residue separately on the basis of T_1-, T_2-, and NOE-based error functions.

2.4.3 Cofactor assay

Fluid-phase cofactor assays were used to measure the cofactor activity of FI-mediated cleavage of C3b to iC3b. Two assays were used; a concentration-dependent assay [151] and a fluorescence assay [152]. The stock proteins for both FI and C3b were purchased from Complement Technologies (Texas, USA).

Concentration-dependent co-factor assay

The concentration-dependent co-factor assay was performed in a final reaction volume of 20 µl in a micro-centrifuge tube (Eppendorf). In each reaction, 3 µg (850 nM in 20 µl (3 µl of stock in 20 µl reaction)) of C3b and 0.05 µg (25.5 nM, 5 µl of 1:10 dilution of stock in 20 µl reaction) of FI were mixed and then FH 1 - 4 at final concentrations of between 0 and 300 nM (see chapter 3, section 3.2 for specific details) was added. The FH samples were prepared from 1 µM stocks. The final reaction volume of 20 µl was reached by adding PBS. The negative control contained the same amounts of C3b and FI but lacked FH. The reactants were mixed and incubated in a circulating water bath at 37 °C (see chapter 3, section 3.2 for specific details). The addition of FH to the reaction was used as the starting point of the reaction. Addition of SDS buffer containing reducing agent (5 µl 4x NuPAGE LDS sample buffer and 2 µL 10x NuPAGE sample-reducing agent), and heating the reaction tube at 95 °C for two minutes, terminated the reaction prior to analysis by SDS-PAGE.

Fluorescence co-factor assay

The fluorescence assay employed in this study for the measurement of fluid-phase FH co-factor activity monitored, in real time, the loss of 8-anilino-1-naphthalenesulfonic acid (ANS) (Sigma Aldrich) fluorescence as C3b was converted to iC3b by FI and FH [152]. Data was collected using a SpectraMax M5 multimode plate reader (Molecular Devices). Excitation was set at 386 nm with emission recorded at 472 nm with a 420 nm cutoff. Stock solutions were pre-warmed to 37 °C. The assay was performed at 37 °C and in PBS, within the wells of a 96-well black polystyrene microplate (Greiner Bio-One) using a reaction volume of 50 µl. Each reaction contained final concentrations of 1.5 µM C3b and 0.05 µM FI to which was added FH to give a concentration of between 0.0125 µM and 8 µM. The reaction was then mixed after the addition of FI and the loss of ANS fluorescence recorded for 5 minutes.

2.4.4 Dynamic light scattering

Dynamic light scattering measurements were performed on a Zetasizer Auto Plate Sampler (Malvern). Protein concentrations were in the range of 0.5-1.0 mg/ml. Samples were filtered (0.2 µm Millipore) prior to use and then 100 µl was loaded into each well of a 384-well plate (Thermo Scientific).

2.4.5 SAXS data collection and analysis

Synchrotron radiation X-ray scattering data were collected at the SAXS/WAXS beam line of the Australian Synchrotron using a 1M PILATUS pixel detector (Dectris, Switzerland). Samples (50 µl) were flowed through a temperature controlled capillary (1.5 mm i.d.) and ~ 10 frames of 2-s exposure time collected. Samples were analysed at 20°C, using protein concentrations of 0.4 – 1.8 mg/ml. The sample-to-detector distance was 1.6 m, covering a range of momentum transfer $0.01 \leq q \geq 0.5$ A^{-1} (where $q = 4\pi\sin(\theta)/\lambda$, with 2θ being the scattering angle and λ=1.13 A the X-ray wavelength). Based on comparison of successive 2-s frames, frames in which radiation damage was detected as an increase in low-q scattering intensity were excluded from averaging.

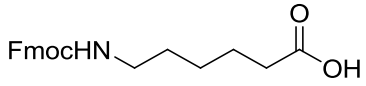
Data from the detectors were placed on an absolute scale (referenced to the known scattering of pure water, 0.016 cm-1), normalized to the transmitted beam intensity, averaged, and the scattering of buffer solutions subtracted. The difference curves were scaled for solute concentration. Data manipulations were performed using the ATSAS software package utilizing the recently developed DATTOOL library for averaging, merging and extrapolation to infinite dilution from concentration series. The forward scattering *I(0)* and radius of gyration, R_g were determined from Guinier analysis, assuming that at very small angles ($s \leq 1.3/R_g$) the intensity is represented as $I(s)=I(0)exp((sR_g)2/3)$. These parameters were also estimated from the full scattering curves using the indirect Fourier transform method implemented in the program GNOM, along with the distance distribution function *p(r)* and the maximum particle dimensions D_{max}. Molecular masses (*MM*s) of solutes were estimated from SAXS data by comparing the extrapolated forward scattering with

that of the water reference. The data was collected and analyzed by Dr Haydyn Mertens.

2.5 Labelling of proteins for EPR

2.5.1 Synthesis of 4-(6-amino-hexaneamido)-2,2,6,6-tetramethyl-piperidyl-1-oxyl for enzymatic labelling

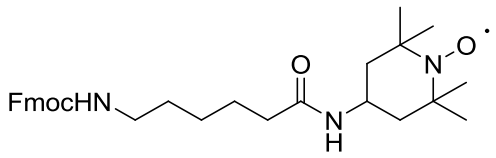
First Step: preparation of 6-(Fmoc-amino) hexanoic acid



A solution of sodium carbonate (1.695 g, 16 equivalents) and 6-aminohexanoic acid (1.048 g) in H_2O (11 ml) and dioxane (7 ml) at 0 °C was stirred for one hour. Fluroenylmethyloxycarbonyl (Fmoc) chloride (2.069 g, 8 mM) was dissolved in 10 ml dioxane and added drop-wise to the mixture and stirred at 0 °C for one hour then at room temperature for 18 hours. The reaction mixture (reddish brown solution) was then poured into 100 ml cold H_2O, mixed and extracted with diethyl ether (3x 50 ml). The aqueous phase was adjusted to pH 2 - 3 and separated with ethyl acetate (3x 50 ml). The organic layer was then dried (magnesium sulfate) to remove water, filtered to remove solid and then dried under rotary evaporation to give a reddish brown solid (2.2 g).

R_f (50% EtOAc in DCM) = 0.50; **mp** 108-109 °C; **IR** (KBr) (cm^{-1}) paper = 3310, 2920, 2850, 1700 sample = 3344, 2942, 1686. **m/z** expected 353.16 (100%), 354.17 (23.1%), 355.17 (3.4%) found 376.1 which is the product plus 1 sodium molecule; **^{1}H NMR** δ (360 MHz, CDC13) 7.70-7.19 (11H ddtts Ar*H*), 4.72 (s, OH), 4.39-4.42 (2H, q, *J* = 6.8 and 7.5 Hz), 4.06 (1H, s,), 3.63-2.98 (2H, qs, *J* = 6.5 and 6.4 Hz, C*H*CH_3), 2.30-2.26 (2H, t, *J* = 7.3 Hz), 1.62-1.54 (2H, q, *J* = 7.2, 7.5, 7.4 and 7.2 Hz), 1.42-1.41 (2H, q, *J* = 6.7 Hz,) 1.31-1.29 (2H, t, *J* = 6.7 Hz), 1.21-1.17 (2H, q, J = 1.4, 5.7, 7.1 Hz); **^{13}C NMR** δ (90.5 MHz, CDC13) 178.56 (C), 156.34 (C), 143.86 (2 x C), 141.21 (2 x C), 127.54 (CH), 126.91 (2 × CH), 124.90 (CH), 119.84 (2 × CH), 47.17 (CH), 40.66 (CH), 33.61 (CH), 29.47 (CH), 25.98 (CH), 24.12 (CH).

Second step: preparation of 6-(Fmoc-amino-hexaneamido)-2,2,6,6-tetramethyl-piperidyl-1-oxyl

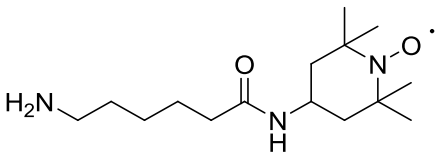


The 6-(Fmoc-amino) hexanoic acid was dissolved in 10 ml tetrahydrofuran (THF) and left to stir in an ice bath. Dry triethanolamine (TEA) (1 ml, 1.1 mM) was added and stirred for 30 minutes at 0 °C. 4-Amino-2,2,6,6-tetramethylpiperidinyloxy **(**4-amino TEMPO) (1.066 g, 1 mM) was dissolved in 8 ml THF and added to the FMOC solution and stirred at room temperature for one hour. To remove the THF the reaction was boiled under reflux for 20 minutes at 70 °C, then the remaining solvents were removed by rotary evaporation to give a reddish brown solid. The compound was dissolved in dichloromethane (DCM) and washed with 0.1 M HCl (2 x 50 ml) and then saturated sodium bicarbonate (1 x 50 ml). Water was again removed using magnesium sulfate and the DCM was removed using rotary evaporation to give a reddish brown semi-solid.

The semi-solid was then chromatographed on Silica gel 60 and eluted using ethyl acetate. The fractions from the brown band were combined and rotary evaporated to remove solvent. Residue was again dissolved in DCM and applied to the fresh silica gel 60 (Sigma Aldrich) column. Once solvent had been removed the result was a reddish brown semi-solid of approximately 2 g.

R_f (50% EtOAc in DCM) = 0.33; **mp** not collected due to product being semi solid; **IR** (KBr) (cm^{-1}) paper = 3320, 3050, 2925, 2850, 1690, 1640, 1600 sample = 2936, 1699, 1652. **m/z** expected 506 (100%), 507.31 (33.1%), 508.31 (6.1%), 507.30 (1.1%) found 507.2; **^{1}H NMR** δ and **^{13}C NMR** δ not collected due to free radical in product.

Third step: preparation of 4-(6-amino-hexaneamido)-2,2,6,6-tetramethyl-piperidyl-1-oxyl



To remove the FMOC protecting group, the residue was first dissolved in DCM and then 20% (w/v) piperidine in dimethylformamide (DMF) was added and left to stir at room temperature for approximately one hour. The solution was then chromatographed on silica gel 60 and eluted with 10% (v/v) methanol in DCM. Fractions containing the product were combined and then rotary evaporated to give a reddish brown semi-solid of 320 mg.

R_f (20% methanol in DCM) = 0.15; **mp** not collected due to product being semi solid; **IR** (KBr) (cm^{-1}) paper = 3400, 3280, 2920, 2850, 1650, 1530 sample = 2940, 1652, 1541. **m/z** expected 284.23 (100%), 28.24 (16.6%), 286.24 (1.7%), 285.23 (1.1%) found 285.2; **^{1}H NMR** δ and **^{13}C NMR** δ not collected due to free radical in product.

2.5.2 Labeling of free cysteine

Labeling of proteins which contained a free cysteine was carried out by incubating the proteins with excess MTSSL spin label (10-20 fold excess) for 4 hours at room temperature in the dark. Un-reacted spin label was removed by excessive dialysis into 20 mM Tris-HCl, pH 7.5.

2.5.3 Labeling of Transglutaminase tag

Enzymatic labeling of the transglutaminase tag was achieved by incubating the proteins (in 20 mM Tris-HCl, pH 7.5) with a 20-fold excess of the synthesized spin label in the presence of 10 mM $CaCl_2$ and 2 µM guinea-pig liver transglutaminase (from lyophilized powder, Sigma Aldrich) at room temperature overnight in the dark. Removal of the un-reacted spin label and transglutaminase from the samples was achieved by size-exclusion chromatography using a Superdex 75 HR 10/30 column in 20 mM Tris-HCl, pH 7.5.

2.5.4 EPR data collection

Samples were dialyzed into deuterated 20 mM Tris-HCl, pH 7.5 and deuterated glycerol was added to each to a final concentration of 50% prior to data collection. Data was recorded at the University of St Andrews using a Bruker E580 EPR machine running pulsed X-band EPR set at 50 K. Analysis of the data was performed using the DeerAnalysis2011 software. Data was collected and analyzed by Dr David Norman, University of Dundee.

2.6 Surface Plasmon Resonance

2.6.1 Binding affinity of FH linker mutants to the bacterial protein PspCN

The binding affinity of the FH linker mutants to PspCN was monitored using surface plasmon resonance (SPR) on a BIAcore T200 sensitivity-enhanced instrument (GE Healthcare). All experiments were carried out at 25 °C using $HBSP^+$ (10 mM HEPES, 150 mM NaCl and 0.05% surfactant P20) supplemented with 50 µM EDTA. The surface of the NTA chip (matrix: carboxymethylated dextran pre-immobilized with nitrilotriacetic acid (NTA), Series S Sensor Chip NTA from GE Healthcare) was prepared by first coating the surface with nickel chloride at 5 µl/min for 25 seconds until ~40-60 RUs were observed. Second, 15 nM hexaHis-tagged SUMO-PspCN (prepared by Dr Andrew Herbert) was flowed over at 30 µl/min for 25 seconds until ~120 RUs was achieved. The reference surface was prepared by only coating it with nickel chloride until the desired amount of RUs was achieved.

The proteins were injected at a flow rate of 100 µl/min for 180 seconds (contact time) followed by a dissociation time of 1500 seconds. For the FH 8-15 linker mutants, single injections were performed for concentration 0, 1, 2, 4, 8, 16 and 32 nM with a double injection performed for 64 nM. For the full-length FH constructs single injections were performed for 0, 0.25, 0.5, 1, 2 and 4 nM with a double injection for 8 nM.

After each injection the chip was stripped (because of the very slow off-rates for FH:PspCN) using 375 nM EDTA with 1 M NaCl followed by 50 mM NaOH and

finally 0.5% (w/v) SDS. All stripping steps were carried out at 30 µl/min for 30 seconds.

Dissociation constants were calculated by fitting kinetic parameters to a one:one Langmuir binding model.

2.6.2 Interaction of full length FH variants with C3b, C3c and C3d

Analysis of the interaction between the full-length FH linker constructs and C3b, C3c and C3d (Complement Technology, Texas) was also performed using the BIAcore T200 instrument. The surfaces of a C1 chip (Series S Sensor Chip C1 with a matrix free carboxymethylated surface, GE Healthcare) were prepared by immobilizing 272 RUs of C3b, 166 RUs of C3c and 69 RUs of C3d using standard amine coupling (BIAcore manual). All experiments were performed at 25 °C using $HBSP^{+}$.

For FH (wild-type) and for the FH (3xGLY) mutant, concentrations of 0, 0.0313, 0.625, 0.125, 0.25, 0.5 and 1 µM were injected singularly whilst a double injection was performed for 2 µM. For the FH (short-linker) mutants (due to sample availability) only single injections were performed for the concentration range 0, 0.0313, 0.625, 0.125, 0.25, 0.5 and 1 µM. A contact time of 90 seconds was used with a dissociation time of 400 seconds at a flow rate of 30 µl/min. Between sample injections the chip was regenerated using two injections of 1 M NaCl for 30 seconds at a flow rate of 30 µl/min.

These experiments were then repeated with the only difference being that the wild-type and mutant versions of FH were pre-incubated with a two-fold molar excess of PspCN. An excess of PspCN was used in an effort to ensure that nearly all the FH had engaged in forming the complex. PspCN had previously been shown not to bind to C3b or its fragments. The concentrations used in these experiments were identical to those used for studying wild-type and mutants versions of FH alone.

2.6.3 Decay-acceleration assay of full-length FH mutants

Analysis of the ability of full-length FH mutants to accelerate the decay of the C3 convertase (C3bBb) was also performed using the BIAcore T200 instrument and a

running buffer of $HBSP^+$ supplemented with 1 mM $MgCl_2$. Using standard amine coupling (see above), 500 RUs of a freshly thawed C3b solution were coupled to a C1 chip. A mixture of 0.5 µM FB and 50 nM FD was then flown over the surface, at a flow rate of 10 µL/min for 120 seconds, to form the C3 convertase (monitored by RUs, achieved was 150 RUs) Subsequently, natural decay of the convertase was allowed to occur for 200 seconds followed by injections of the FH variants for 120 seconds at 10 µL/min. Concentrations of wild-type FH and FH mutants used were 2.5, 5 and 10 nM. The surface was regenerated between injections with 30 sec of 0.1 µM FH+0.2 µM PspCN at 10 µl/min followed by 30 sec 1M NaCl at 10 µl/min.

The experiment was repeated except that (as before) FH was pre-incubated with PspCN (2:1 ratio of PspCN:FH) . The concentrations used in these experiments were identical to those used for studying wild-type and mutants versions of FH alone.

Chapter 3: Study of disease-associated sequence variations in CCP 1 of FH

Various mutations and polymorphisms found within the genes encoding some proteins (CR1, MCP, FI, FB, C3 and FH) of the complement system have been linked with aHUS, DDD and AMD. The main pathogenic feature of these three diseases is dysregulation and/or excessive activation of the AP [153-155]. As described in more detail in the Introduction, FH has at least two common SNPs that are linked to the risk of developing both AMD and DDD: Y402H is part of a haplotype that increases risk (six-fold for homozygous individuals and 2.5 times for heterozygous individuals) [107] while I62V is part of another common haplotype that is protective for DDD and AMD [95].

Background information for study.

As also described in the Introduction, the majority of mutations linked to aHUS and found in FH are heterozygous single-point mis-sense mutations found towards the C-terminus and especially in CCPs 19 and 20 [156, 157]. It has been hypothesized that these mutations may lead to a dominant-negative effect. A study by Heinen *et al* [158] showed that the FH mutant E1172STOP (CCP 20 deletion) purified from a heterozygous patient's plasma showed defective binding to C3b, C3d, heparin and the surface of endothelial cells. When regulatory activities were compared, the mutant protein was as active as wild-type FH in fluid phase but on the cell surface activity was reduced. The patient's plasma was shown to contain both wild type and the mutant protein at almost equal levels and when compared to wild-type FH (from a healthy control) the heterozygous mixture displayed reduced binding (six-fold reduction in mean fluorescence detected) to human umbilical vein endothelial cells. The reduction in the cell-binding suggests an interaction between the mutant and wild-type protein, which results in the wild-type protein function being affected. This reduction in activity for wild type FH could indicate the operation of a dominant negative effect.

When in solution FH is believed to exist as 95% monomer [159] but on the cell surface FH may well behave differently. It has been suggested that FH forms oligomers through its C-terminus [64] and the oligomers can bind to the cell surface either through GAGs [160] or C3d [161]. The Heinen study showed that, after treating both proteins with the cross linker DMA and analysis using western blot, both wild-type FH and mutant FH could form dimers but the wild-type protein seemed to form a dimer more readily with itself.

As discussed in the Introduction, the N-terminal region containing the first four CCPs of FH has three important roles; it competes with FB for binding to C3b; it acts as a vital cofactor for the FI-mediated cleavage of fluid-phase C3b to iC3b; and it accelerates the decay of C3b.Bb, the C3 convertase complex that cleaves C3 to C3a and C3b. Interestingly the N-terminal module (CCP 1) contains two mutations linked to aHUS, namely R53H and R78G, as well as the AMD/DDD-protective SNP I62V [93, 94] mentioned above. Work performed on R53H and R78G showed that both disrupt the cofactor function of FH but do so by different mechanisms [53]; the former blocks interactions with FI while the latter blocks interactions with C3b. Both also disrupt decay-accelerating activity, which proceeds via an unknown mechanism.

The work in this chapter aimed to investigate one possible mechanism of a putative dominant-negative effect for the two aHUS-linked CCP 1 mutants (and the I62V SNP) by assessing the extent to which the mutant proteins interfered with the cofactor activity of the wild-type protein. Mutations were made in the context of the recombinant, four-CCP module, N-terminal truncation fragment FH 1-4. These proteins were produced and purified from *P. pastoris*. Fluid-phase cofactor assays were developed that allowed both "homozygous" (either wild-type or mutant proteins, alone) and "heterozygous" (a mixture of wild-type and mutant FH 1-4) conditions to be compared.

3.1 Production and purification of FH 1-4 proteins.

A clone containing the DNA for each of three FH 1-4 constructs - R53H (on a V62 background), I62V and R78G (on a V62 background) - in the pPICZαB expression vector had been prepared previously [53]. Each of the constructs contained an N-

terminal Myc tag (EQKLISEEDL) and a C-terminal *hexa*-His tag to aid purification and reduce handling time, and because FH 1-4 without tags was found to be susceptible to proteolysis. A sample of FH 1-4 V62 had been prepared previously by Dr Christoph Schmidt in the lab.

3.1.1. Small-scale tests of protein production

After each vector containing the DNA encoding the FH 1-4 variants had been transformed successfully into the KM71H strain of *P. pastoris*, between two and four colonies were chosen for small-scale tests of protein production (see Methods Section 2.2.4). The supernatant from each culture was applied manually to a 1 ml HisTrap chromatography column and eluted using an increasing gradient of imidazole (to 500 mM, see Methods Section 2.3.2). The results of these test productions were assessed using gradient SDS-PAGE (Figure 3.1).

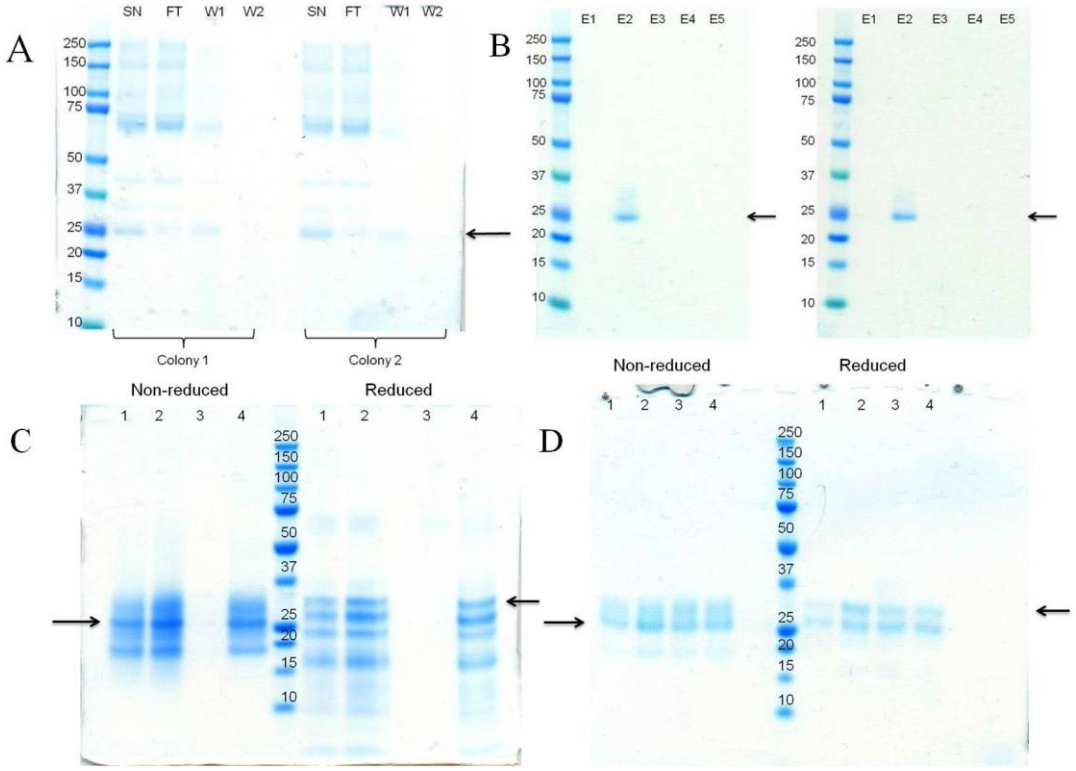


Figure 3.1 Small-scale protein production trials for disease-associated sequence variants of FH 1-4. In (A) the gel (run under non-reducing conditions) shows, for each of two colonies, the supernatant from I62V-production trials before application to the HisTrap column (SN); the flow through (FT); and fractions (W1 and W2) collected from two sequential wash-steps applied before commencing the imidazole gradient. The gel in (B) (run under non-reducing conditions) shows, for each I62V colony, the five elution fractions (E1-E5) from a HisTrap column as in A; protein of the

expected mass was detected exclusively in fraction E2. Production trials for R53H are shown in (C) and for R78G in (D). For these gels, the lane numbers indicate the colony number; only the second elution fraction was run on the gel in each case. Samples for (C) and (D) were run under both reducing and non-reducing conditions as indicated above each gel. The black arrows indicate the expected migration position, under non-reducing and reducing conditions.

These protein-production trials showed that both I62V colonies produced a species that migrated in the position expected for non-reduced FH 1-4 (at about 25 kDa). In the case of R53H, colonies 1, 2 and 4 all produced correctly sized protein as did all four R78G colonies; but the bands were weak compared to those obtained for I62V while for R53H in particular, substantial degradation was apparent under reducing conditions. The next stage was to perform a larger-scale protein production trial, in 2 L shaking flasks, for the three variants and it was decided to use colony 2 for inoculation in each case. In these trials, as described in Methods, both the BMG and BMM media were enriched with 10% (w/v) cas-amino acids - these media will henceforth be referred to as BMG-CA and BMM-CA.

3.1.2 Larger-scale production trials of FH 1-4 R53H in shaking flasks

An initial volume of 5 L BMG-CA was used for cell culture as described in Methods Section 2.2.5. Cells were subsequently spun down and re-suspended in 1.25 L of BMM-CA in ten 2 L baffled flasks (*i.e.* 125 ml per flask) for induction (see section 2.2.5). After harvesting, the supernatant was adjusted to pH 7.5, filtered (0.22 µm) and applied to a 25 ml XK 16/20 IMAC chromatography column (GE Healthcare) charged with Ni^{2+}. The elution was performed using a three-step imidazole gradient (to 1 M, see Methods Section 2.3.2). The resulting chromatogram is shown in Figure 3.2 alongside SDS-PAGE analysis of the fractions.

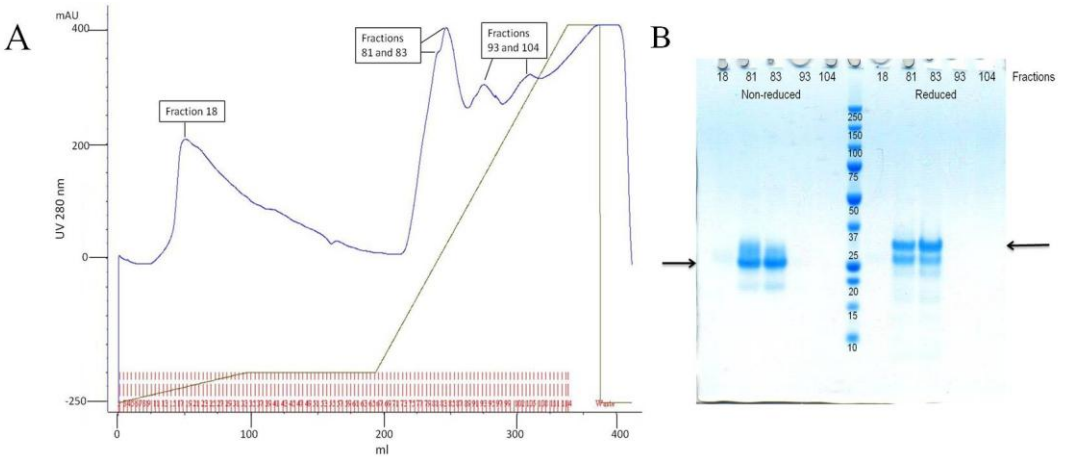


Figure 3.2 R53H eluted from Ni^{2+}-affinity resin. (A) Shows the chromatogram ($OD^{280\ nm}$, blue line; black line shows imidazole gradient) for R53H following Ni^{2+}-affinity chromatography. Four main peaks are present. In (B) are the results of gradient SDS-PAGE for the fractions shown in A, run under both reducing and non-reducing conditions as indicated. Proteins migrating to the expected positions on the gel (indicated by black arrows) were found in fractions 81-83.

The elution profile from the Ni^{2+}-affinity column contained four main peaks. SDS-PAGE analysis of the fractions showed that despite the detection of material absorbing at 278 nm, no or very little protein was present in fractions 18, 93 or 104. In fractions 81 and 83, on the other hand, three bands were detected under non-reducing conditions; the dominant band migrating just above the 25-kD marker is likely to be R53H FH 1-4 and is accompanied by a smeary band above, and a very faint band below. While the lower band is most likely a degradation product of R53H FH 1-4, the origin of the higher smear is unknown. Since an N-glycosylation site is present in CCP 4 it was initially hypothesized that this had become glycosylated, but following treatment with an endoglycosidase enzyme the smear did not improve (not shown). Under reducing conditions, two main bands are detected, with the lower band likely corresponding to FH 1-4 that had been proteolytically clipped (but remains intact until the disulfides have been reduced). The upper band under reducing conditions – running a little more slowly than the same material under non-reducing conditions – is almost certainly the FH 1-4 mutant. The next purification step was performed after concentration and buffer-exchange into PBS. This consisted of size-exclusion chromatography using a Superdex 75 HiLoad 16/60 column; two separate injections were performed (see below).

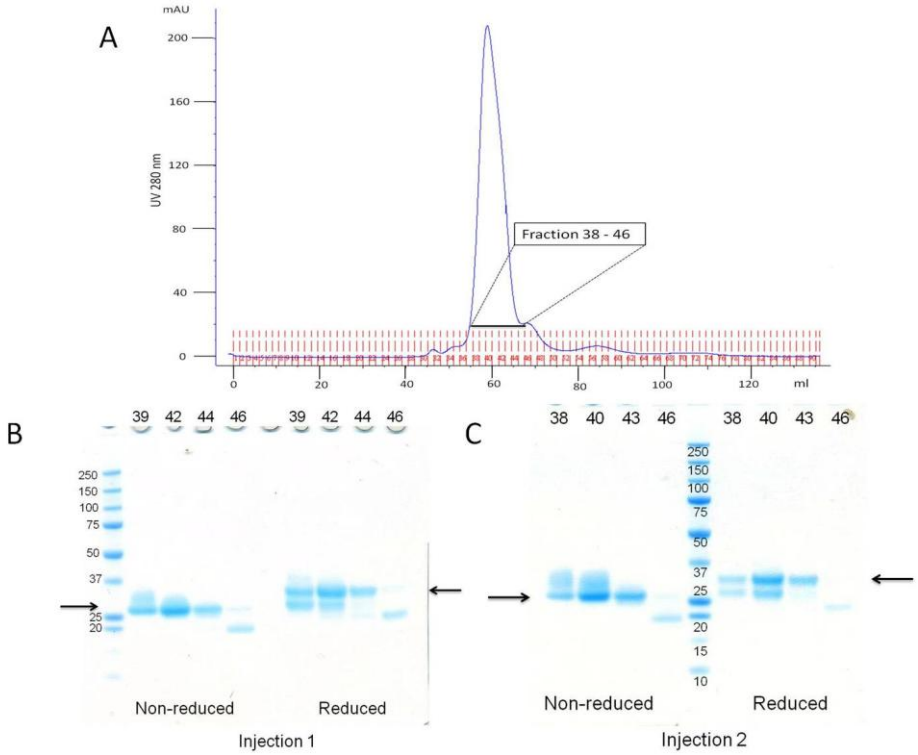


Figure 3.3 Size-exclusion chromatography of R53H FH 1-4. The size-exclusion chromatogram shown in (A) is from the second injection. For the gradient SDS-PAGE gels in (B) the lane numbers indicate the fractions numbers (also indicated in (A)) for both injections; black arrows indicate the expected migration positions for FH 1-4 under non-reducing and reducing conditions.

The size-exclusion chromatogram for injection 2 contained one main protein peak eluting at 60 ml. SDS-PAGE showed that the smear is still present in this peak, and under reducing conditions there appears to still be some clipped protein. Across the peak there is some resolution of these species, so it was decided in the interest of yield to pool (separately) protein fractions 43, 44 and 45 from the two injections. These were run on an SDS-PAGE gel in non-reducing and reducing conditions (Figure 3.4), and on this bases, fractions from 43 and 44 were judged to be the most pure out of those tested, and the material they contained was used in the cofactor assays.

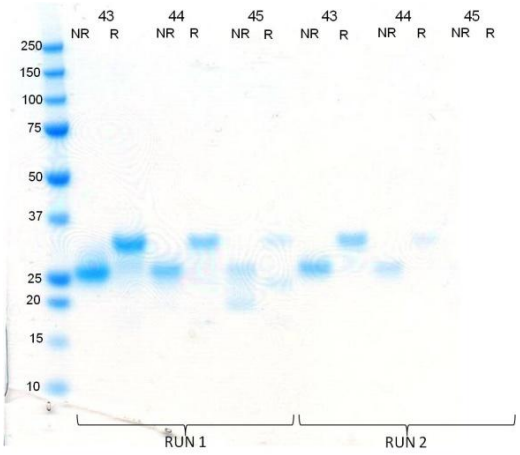


Figure 3.4 Final quality check for R53H FH 1-4. The gradient SDS-PAGE gel shows fractions 43 – 45 (pooled from two runs) under non-reducing (NR) and reducing (R) conditions.

The concentration of each fraction in a final volume of 1.5 ml is shown in Table 3.1

Table 3.1 Concentrations of R53H fractions.

Run	Fraction	Concentration (µM)
1	43	16
	44	10
	45	4.5
2	43	11
	44	6

The fractions from run 1 were combined and gave a concentration of 43.7 µM (1.2 mg/ml) in a final volume of 1 ml (from 1.25 L of BMM-CA final concentration was ~ 1.3 mg/L).

3.1.3 Larger-scale production trials of FH 1-4 I62V in shaking flasks

Using the same method as employed for R53H FH 1-4, a starting volume of 5 L BMG-CA media was used and then cells transferred for induction to 1.25 L BMM-CA media. As before, the supernatant was adjusted to pH 7.5, filtered and applied to the 25 ml XK 16/20 IMAC column charged with Ni^{2+} (Figure 3.5).

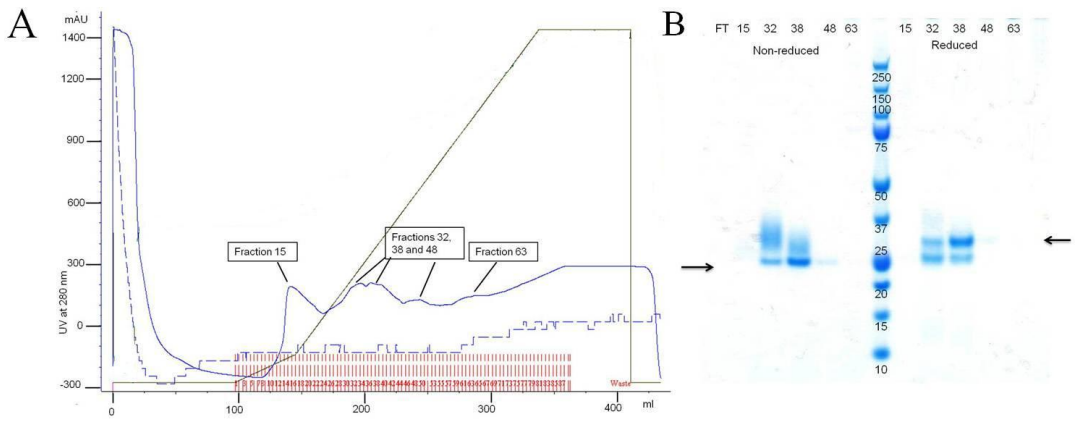


Figure 3.5 Ni^{2+}-affinity chromatography and SDS-PAGE gel for I62V FH 1-4. The Ni^{2+}-affinity elution profile for I62V (A) shows multiple peaks. The constant blue line shows absorbance at 280 nm and the dark green line shows the imidazole gradient. In (B) the flow-through (FT) was run to check for protein that did not adhere to the column, while the numbered lanes correspond to the fraction number analyzed. The black arrows in (B) indicate the expected position of FH 1-4 in non-reducing and reducing conditions.

Analysis by SDS-PAGE shows that a species likely corresponding to I62V FH 1-4 was present in fractions 32 and 38 (elution at 200-225 ml). There was no detectable protein in the lane corresponding to the flow-through (FT). As was seen in the case of R53H FH 1-4, a higher-molecular weight smear is present which did not disappear after endoglycosidase treatment. Size-exclusion chromatography was performed on I62V FH 1-4-containing fractions as before (Figure 3.6).

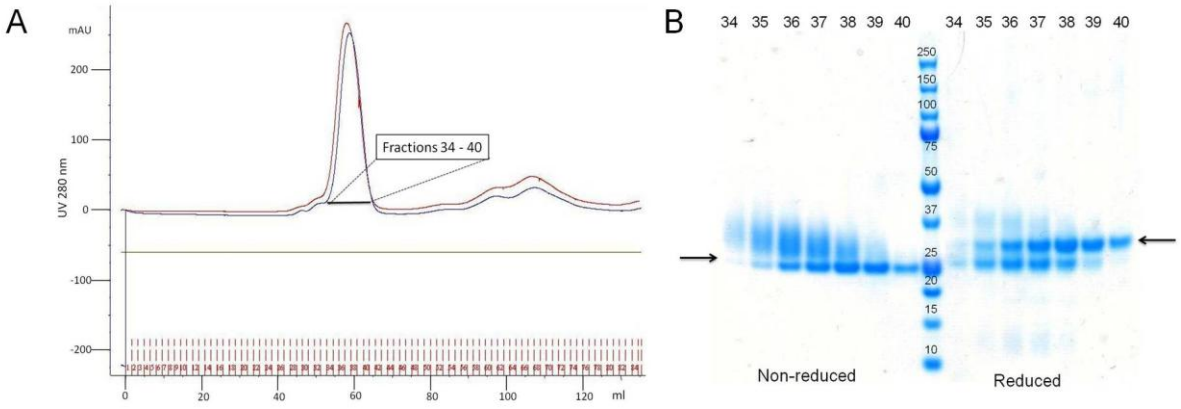


Figure 3.6 I62V FH 1-4 size-exclusion chromatogram and SDS-PAGE. The elution profiles (red and blue) for duplicate injections of I62V FH 11-4 (eluted from the Ni^{2+}-affinity column) are overlaid in (A). In (B) are the results of gradient SDS-PAGE showing fractions 34 – 40 from injection 1 run under both reducing and non-reducing conditions. The expected migration position of FH 1-4 is shown by the black arrows.

Both the higher-molecular weight smear and the clipped material (visible on the gel under reducing conditions) are partially resolved from the target material. Injection 2 showed a similar result, so fractions 39 and 40 were kept from both injections for cofactor analysis. After being combined and concentrated the final sample for I62V was 30 µM (0.8 mg/ml) in 2 ml (from 1.25 L of BMM-CA final concentration was ~ 0.6 mg/L).

3.1.4 Larger-scale production trials of FH 1-4 R78G in shaking flasks

In this case 4 L of BMG-CA media was used initially followed by 1 L of BMM-CA media. As for the other FH 1-4 sequence variants, a XK 16/20 IMAC column charged with Ni^{2} was used for the first step. The elution was performed using a three-step gradient of imidazole (final concentration was 1 M).

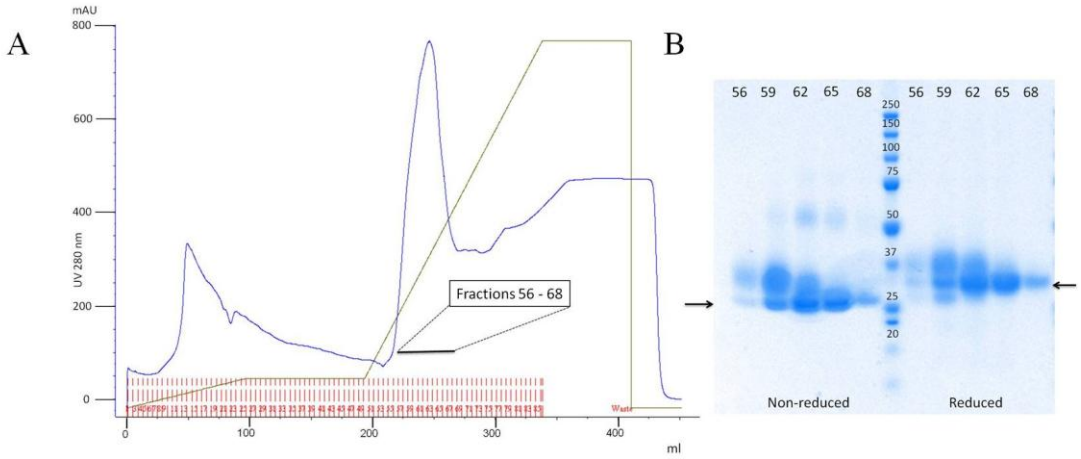


Figure 3.7 Ni^{2+}-affinity elution chromatogram and SDS-PAGE of R78G FH 1-4. The Ni^{2+}-affinity elution profile (blue) is shown in (A) with the dark green line indicating the imidazole gradient. The corresponding SDS-PAGE is shown in (B).

Fractions corresponding to the main peak (elution at 250 ml) of the Ni^{2+}-affinity chromatogram contain species that migrate as expected for the target protein, along with the familiar higher-molecular weight smear. A small amount of dimer (running at just above 50 kDa) may be present in the non-reduced fractions from the middle of the peak. As before fractions 56-68 were pooled and subjected to size-exclusion chromatography (Figure 3.8).

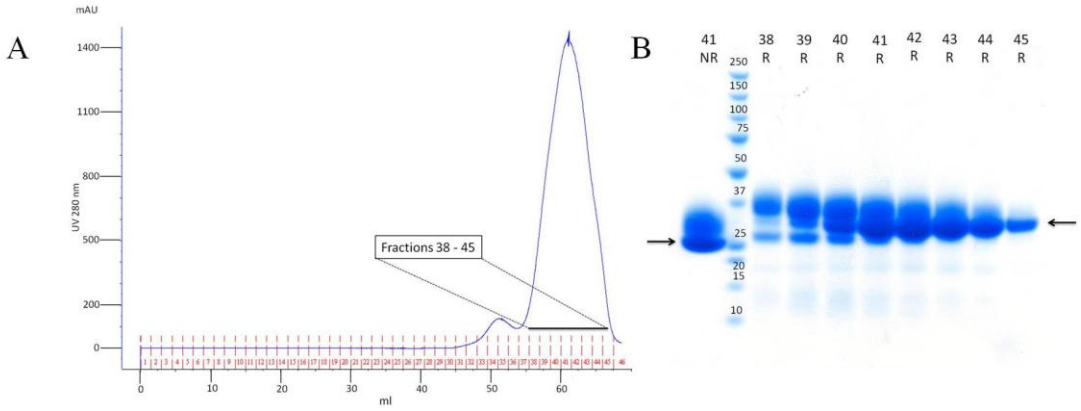


Figure 3.8 Size-exclusion chromatogram of R78G FH 1-4. The elution profile is shown in (A). The fractions indicated in (A) were then run in reducing conditions on SDS-PAGE (B) alongside fraction 41 run under non-reducing conditions. Arrows indicate the expected position on the gel of the target protein.

Analysis by SDS-PAGE shows that the smear reduces towards the end of the peak with very little present in fraction 45. Fractions 38 - 43 were combined, concentrated and exchanged into 20 mM glycine with 70 mM NaCl buffer, pH 9.75, as a prelude to high-resolution anion-exchange chromatography. A gradient to 500 mM salt was run over 20 column volumes, then three column volumes were used to reach 1 M NaCl (Figure 3.9).

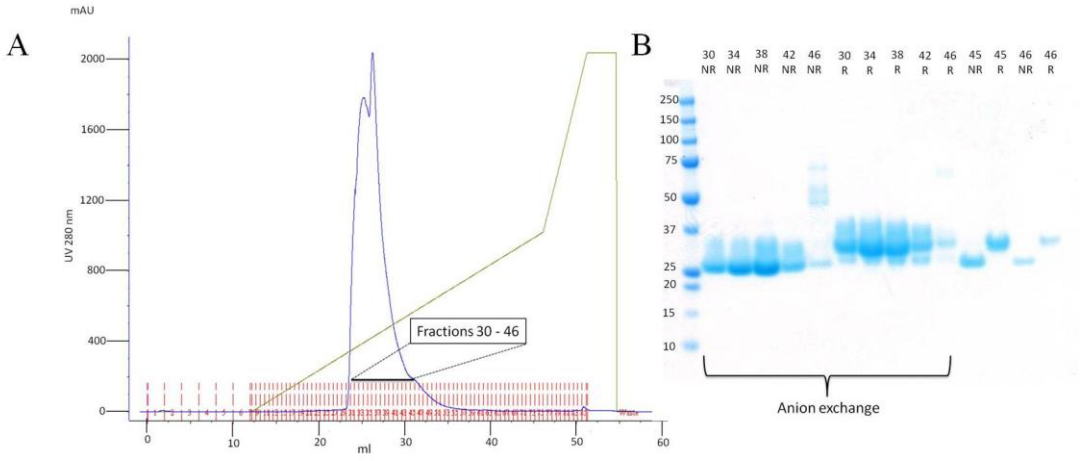


Figure 3.9 Anion-exchange chromatography of R78G. (A) Anion-exchange elution profile (blue), with green line showing NaCl gradient to 1 M final concentration. The SDS-PAGE in (B) shows fractions taken from across the peak in both non-reducing (NR) and reducing (R) conditions. The last four lanes on the gel are fractions 45 and 46 from the size-exclusion step.

Anion-exchange chromatography of R78G resulted in two incompletely resolved peaks and analysis by SDS-PAGE showed that the smear was still present at the

same level as in the size-exclusion fractions. Fractions 30-42 were stored but unused due to the smear. Size exclusion fractions 45 and 46 were run in the last four lanes of the SDS-PAGE under non-reducing and reducing conditions to check the purity and were stored for cofactor analysis. The final concentration when the fractions were combined was 18 µM (0.5 mg/ml) in a volume of 1 ml (from 1 L of BMM-CA final concentration was ~0.5 mg/L).

3.2 Cofactor activity of disease-linked sequence variants.

3.2.1Optimizing FH concentrations[1]

SDS-PAGE cofactor assay

A standard fluid-phase cofactor assay [151] – based on the use of SDS-PAGE to detect the cleaved alpha chain into 67-kDa and 43-kDa fragments - was adapted and optimized in an attempt to test hypotheses regarding the dominant-negative effects of FH sequence variants. It was important to ensure that the assay would be responsive (*i.e.* the outcome would be roughly proportional) to the amount of functional FH present. This would ensure the outcome would be sensitive to the possibility of the (inactive) mutant FH interacting with wild-type FH in such a way as to reduce the latter's activity. With this in mind, a concentration series of FH 1-4 was investigated in the SDS-PAGE-monitored assay as described below.

The concentration range of FH 1-4 used in a previously reported cofactor activity was 3-700 nM [162]. On this basis a dilution series was performed starting with 300 nM FH (stock was 1 µM). The concentration of C3b were kept at 0.85 µM (stock was 7 µM), and the concentration of factor I (FI) at 28.4 nM (stock was 14 µM), in a final volume of 20 µl with an incubation time of 10 minutes at 37 °C. A threefold dilution series was used and thus the final concentrations of FH were; 300 nM, 100 nM, 33 nM, 11 nM and 3 nM. An additional two concentrations of 75 nM and 25 nM were also included in an attempt to detect smaller differences between the mutants

[1] The data from the concentration assay was published in 53 Pechtl IC, Kavanagh D, McIntosh N, Harris CL, Barlow PN. Disease-associated N-terminal Complement Factor H Mutations Perturb Cofactor and Decay-accelerating Activities. Journal of Biological Chemistry 2011, 286: 11082-11090

(Figure 3.10). The FH 1-4 V62 protein used in all these experiments was prepared by Dr Christoph Schmidt.

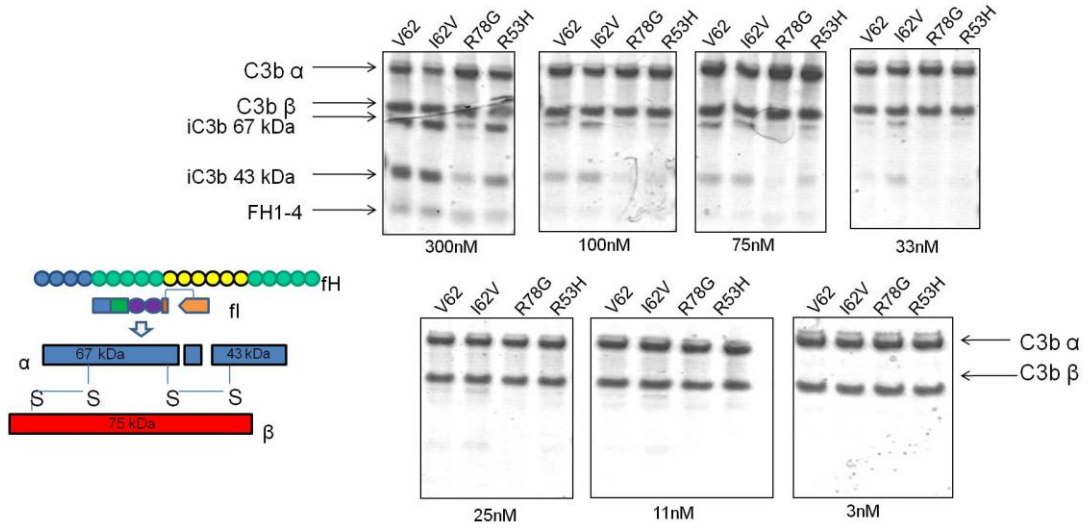


Figure 3.10 Factor H 1-4 concentration-dependent cofactor assay. The concentrations of FH 1-4 used to perform this assay were 300 nM, 100 nM, 75 nM, 33 nM, 25 nM, 11 nM and 3 nM. All four proteins have cofactor ability at 300 nM but R78G is the weakest. At 100 nM, a very clear difference is visible with a decrease in R78G activity, which is no longer detectable at 75 nM. R53H has detectable cofactor activity at 75 nM but not at 33 nM. At the lower concentrations a difference is observed between V62 and I62V in that for 25 nM I62V a 43-KDa band is detected but this is not the case for V62.

This experiment (Figure 3.10) showed that R78G is the least active variant followed by R53H. The comparison between V62 and I62V reveals that I62V has slightly more cofactor ability than V62 (on the basis that more iC3b is present at 25 nM for I62V than V62); this was also seen in the study by Tortajada *et al* [95] where the I62V FH variant (purified from homozygous plasma) was ~20% more active than the V62 variant (also from homozygous plasma) in a fluid-phase C3b assay. At the lowest concentration of 3 nM none of the four proteins had detectable activity over the 10-minute incubation period.

Based on this concentration series, conditions for the test of the dominant-negative hypothesis (see section 3.2.3 for results) were selected. It was desirable to use a 1:1 ratio of wild-type:mutant (on the assumption that heterozygous individuals have an equal amount of each protein) and a concentration of V62 for which activity (*i.e.* the iC3b 43-kDa band) should be detectable in this assay along with a concentration of

R53H or R78G for which no cofactor activity could be detected. Based on this reasoning, 40 nM of each protein was selected.

Fluorescence cofactor assay

To monitor the cofactor ability of the mutants in real time a fluorescence assay was used [152] where the conversion of C3b to iC3b results in a reduction of fluorescence in the presence of a fluorophore, ANS, that binds less well to iC3b [14].

The assay was adapted from the original protocol to use significantly smaller volumes in 96-well plates (see Methods). The main reason for this was to conserve proteins, and in particular C3b. But the use of smaller volumes did increase the risk of errors in consistency and accuracy. Another challenge was to keep the 96-well plate at a constant 37 °C; to ameliorate this problem all proteins as well as the buffer were incubated at 37 °C throughout the experiments. Timing issues were also challenging; after the addition of FI to the ANS:C3b:FH complex there was a delay in recording data due to a mixing step of 5 seconds performed by the plate reader.

This assay had not been performed previously in the lab. In order to test it, both plasma-purified FH (5 μM stock, Complement Technologies) and a recombinant version of full-length FH containing the aHUS-linked mutation, D1119G (6.4 μM stock, prepared by Dr Heather Kerr, Edinburgh) were used. A two-fold dilution series was used with the final concentrations of FH being 200 nM, 100 nM, 50 nM, 25 nM and 12.5 nM. All experiments were performed in duplicate and each reaction mixture contained 1.5 μM C3b 50 nM FI and 100 μM ANS in a final volume of 50 μl PBS; all experiments were performed at 37 °C.

The results for plasma-purified FH (Figure 3.11 (A)) showed a decrease over time in relative fluorescence units (RFU), consistent with C3b being cleaved to iC3b. The initial rate of iC3b production/C3b loss increased with higher concentrations of FH (Figure 3.11 (B)) although the relationship between rate and FH concentration (this will depend on the affinity of FH for C3b and the K_m of the C3b:FH complex for FI) in this set of test assays was unclear. Note that to emulate physiological conditions of FH in plasma, the final concentration would need to be approximately 1-2 μM.

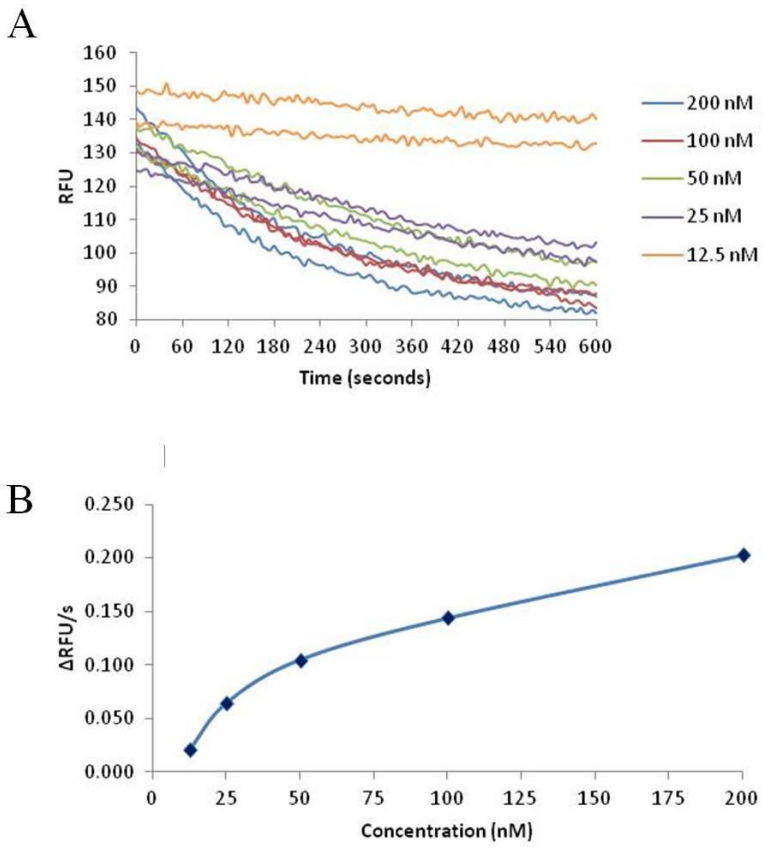


Figure 3.11 Fluorescence-based cofactor assay for plasma-purified FH. (A) Loss of fluorescence over time reflecting conversion of C3b to iC3b. (B) Rate of iC3b production as a function of FH concentration.

When FH containing the D1119G mutation was tested (Figure 3.12 (A)), a plot of ΔRFU/s against FH concentration (Figure 3.12 (B)) showed a linear relationship, which is more in line with expectation, although rate was not proportional to FH concentration. The differences in shape between Figures 3.11(B) and 3.12(B) are probably not significant and likely reflect technical issues of temperature control, timing, and the errors associated with small differences in fluorescence for the lower FH concentrations. It is potentially interesting that the plasma-purified FH appears to be three times more active than the recombinant D119G mutant but this was not further pursued here since our focus was on the CCP 1 mutants.

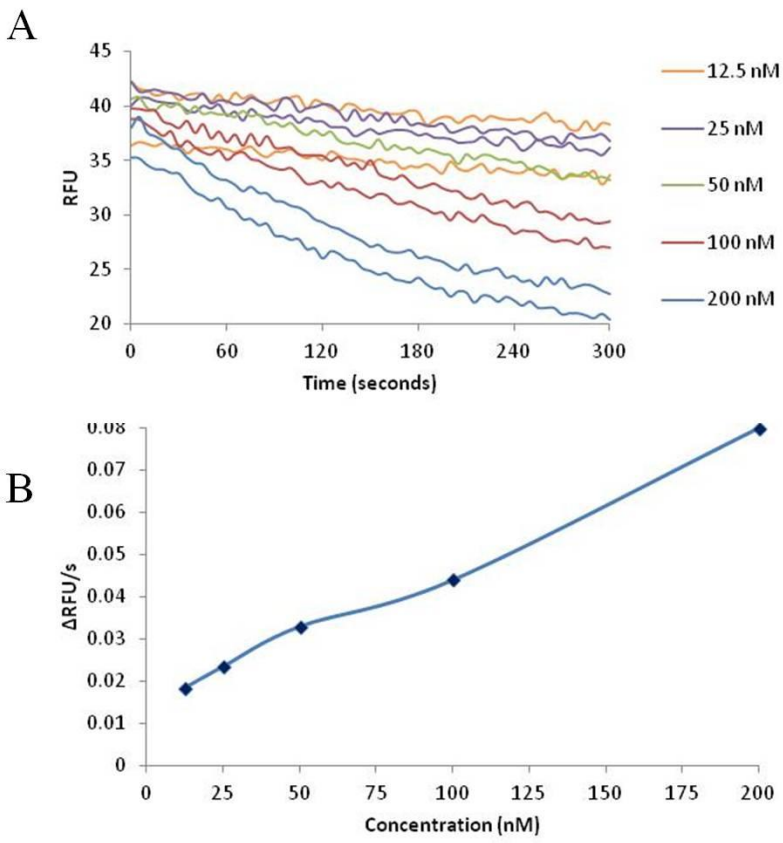


Figure 3.12 Fluorescence cofactor assay for FH containing D1119G mutation. As in Figure 3.11.

Results from these trials with full-length FH were encouraging, so the assay was repeated using identical concentrations of V62 FH 1-4. Little to no conversion was observed so it was concluded that FH 1-4 is not, in this assay, as active as full-length FH in terms of cofactor activity. Subsequently, 400 nM plasma-purified and recombinant FH (D1119G) were compared to 4 µM V62 FH 1-4 (Figure 3.13 (A)). This suggested that the full-length proteins were about ten times more active than FH 1-4. This was an unexpected result – given that the dogma states that CCPs 1-4 are necessary and sufficient for the fluid-phase cofactor activity of FH – but it was not pursued further in this work. While (Figure 3.13 (B)) the full-length D1119G mutant protein appeared less active in this comparison than the plasma-purified protein (as above), this was also not investigated further although again it suggests a role for C-terminal modules in fluid-phase complement regulation.

Based on the above results it was decided to use a two-fold dilution series of V62 FH 1-4 and the variants, with final concentrations of 0.5 µM, 1.0 µM, 2.0 µM, 4 µM and 8 µM.

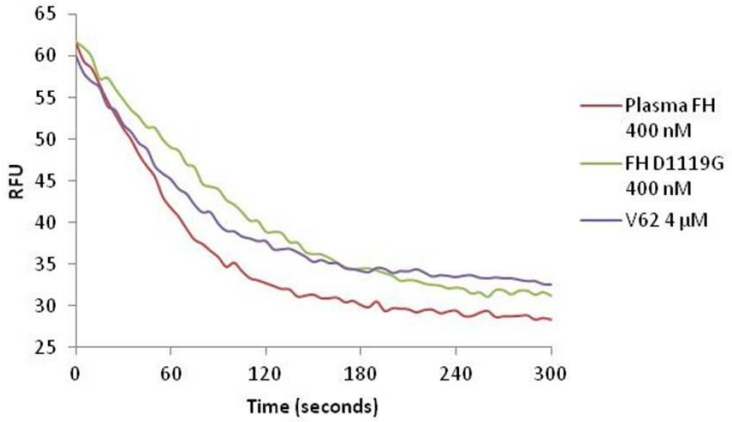


Figure 3.13 Fluorescence cofactor activity assay: full-length FH vs. FH 1-4 V62. The conversion of C3b to iC3b over time for each protein monitored by loss of fluorescence.

The results for fluorescence concentration series (Figure 3.14) were noisy and difficult to interpret. At the highest concentrations of FH 1-4 used (4 and 8 µM), all three proteins exhibit very roughly comparable cofactor activity even though this was well below the K_D reported for R78G (K_D > 35 µM). At the lowest concentration (0.5 µM), none of the FH 1-4 proteins yielded much activity (although the initial reading was very low for V62 so an initial fall in fluorescence may have been missed for technical reasons). The reaction mixtures containing 1 and 2 µM FH 1-4 afforded the best comparisons. In these cases, FH V62 was more active than R78G with R53H being the least active.

Thus the SDS-PAGE and real-time fluorescence assays suggest differing conclusions with regards to the relative activities of the two mutants, but both support the hypothesis that these disease-related mutants are deficient in cofactor activity compared to the wild-type.

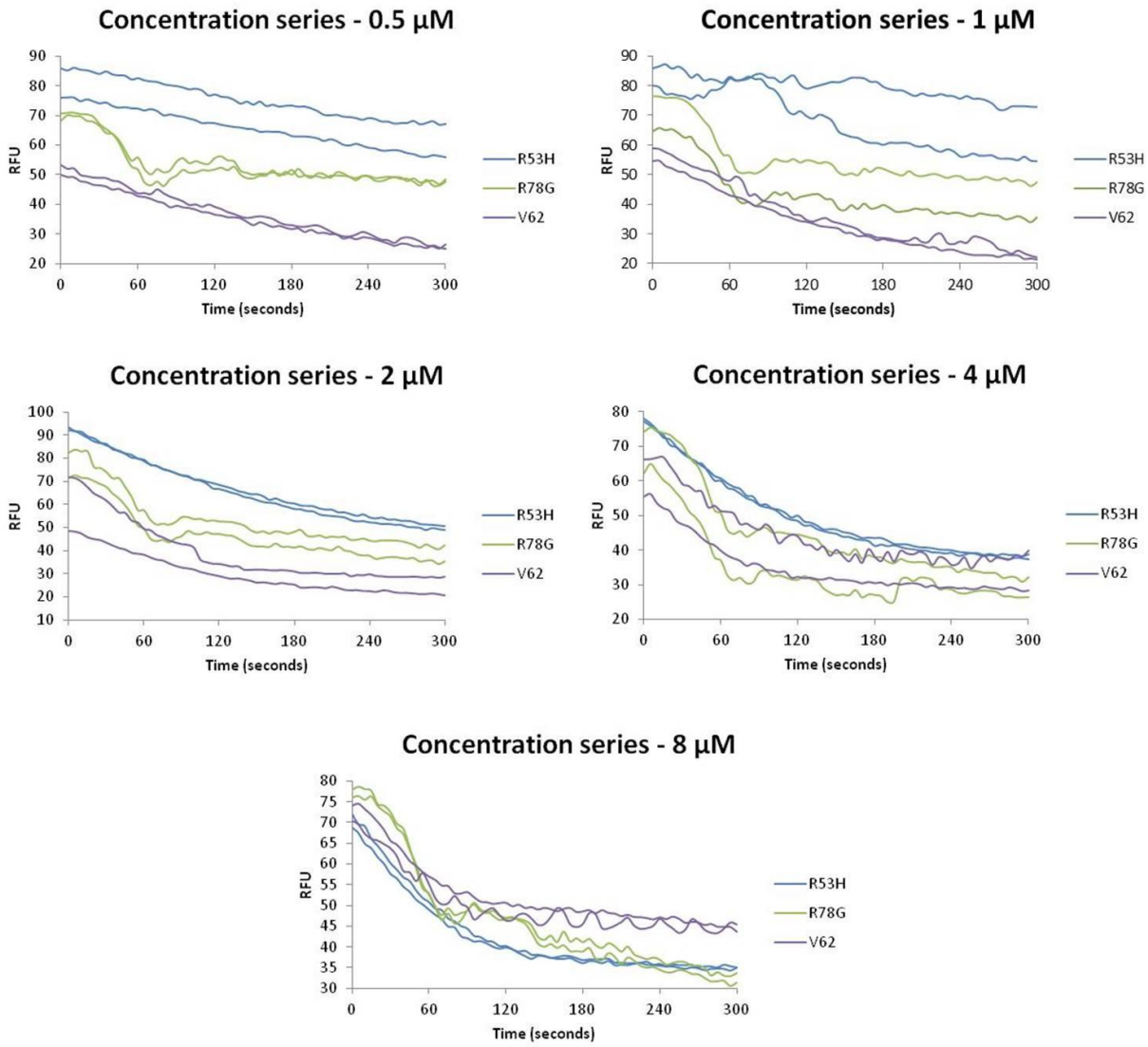


Figure 3.14 Fluorescence cofactor assay – concentration series. Each graph shows the conversion of C3b to iC3b over time for each concentration (0.5, 1, 2, 4 and 8 µM).

3.2.2 Inhibition Assay

The intention was to test if the FH 1-4 fragment containing a disease-linked mutation could inhibit the cofactor activity of the wild-type protein (in this study V62 FH 1-4) by performing SDS-PAGE-based assays on mixtures of the two proteins. While the ratio of wild-type FH to mutant FH may be one-to-one in the plasma of a heterozygote, it was decided for the purposes of this test, to maintain a total concentration of 40 nM FH but to vary the ratio of wild-type and mutant proteins, starting with "40 nM wild-type plus no mutant" and progressing to "no wild-type

plus 40 nM mutant". The concentration of FI was 25.5 nM while C3b was kept at 860 nM (as determined in section 3.2.1 SDS-PAGE assay). To avoid confusion, sample conditions were named using a system based on the concentration of each protein present, *i.e.* 40:0 means 40 nM V62 and 0 nM of the mutants, while 30:10 means 30 nM V62 and 10 nM of the mutants. Each of the experiments was performed in duplicate.

A trial assay was performed using PBS instead of the mutant proteins (over the 0-40 nM wild-type (V62) FH 1-4 range). Inspection of the resulting SDS-PAGE (Figures 3.15 and 3.16) showed no differences in the amount of iC3b produced (judged from densitometry of the bands attributed to the 43-kDa and 67-kDa fragments of the alpha-chain) for FH concentrations between 40 nM and 15 nM FH. Below 15 nM FH, there was an approximately linear relationship between cofactor and iC3b production. Although it might have been preferable to have included (in the subsequent experiments) more FH concentrations that lay within this linear range, the effects of co-incubation with mutants could not be predicted, so it was decided to stick with the 40 nM total concentration of wild-type and mutant.

In the case of R53H, SDS-PAGE (Figures 3.15 and 3.16) showed that there was little difference in iC3b production between 40 nM FH and 15 nM FH + 25 nM mutant FH. The previous study had shown that 25 nM R53H produced no detectable iC3b so nearly all of the iC3b produced in the 15:25 sample must be due to the 15 nM V62 FH 1-4. That the quantity of iC3b produced matches what would be expected from this concentration of V62 FH 1-4 in the absence of any inhibitor implies that the presence of R53H FH 1-4 had an insignificant effect on wild-type (V62) FH 1-4 cofactor activity. At concentrations of V62 FH below 15 nM (and of R53H above 25 nM), cofactor activity drops off in a very similar manner to the trial experiment (no R53H). This reinforces the conclusion that the mutant does not inhibit the "wild type".

A broadly similar result was obtained for R78G (Figure 3.15 and 3.16). As before, iC3b production was essentially constant over the 15-40 nM V62 FH 1-4 (25-0 nM R78G) range but fell off below 15 nM in way that was independent of the presence

or absence of R78G. In conclusion, these experiments demonstrate the absence of an interaction between wild-type (V62) FH 1-4 and either of the mutants under investigation.

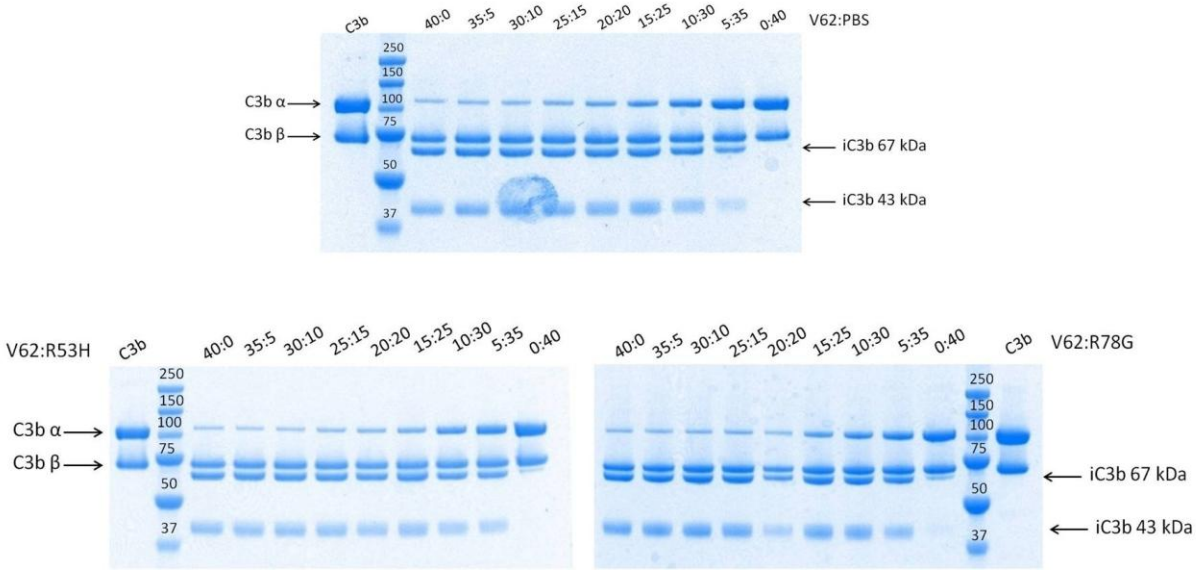


Figure 3.15 Inhibition assay for R53H and R78G. The SDS-PAGE shown are for the PBS trial gel, V62 against R53H and V62 against R78G.

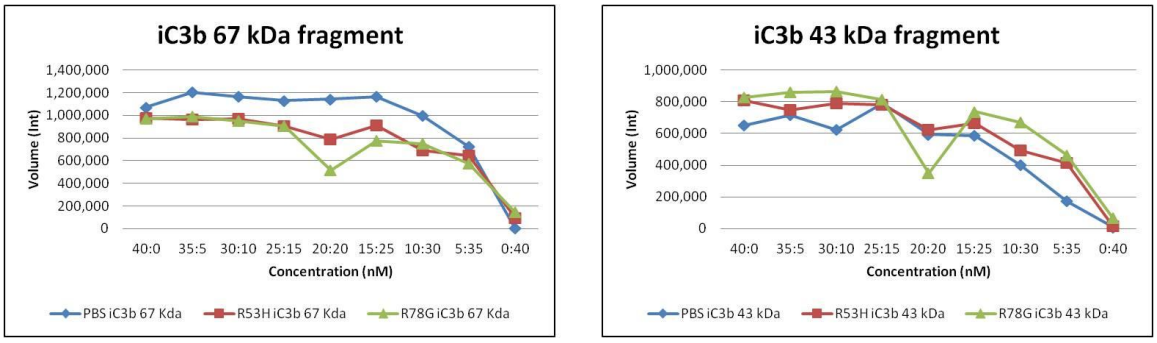


Figure 3.16 Quantification charts for inhibition assay. The left chart shows the average result for the quantification of the iC3b 67 kDa bands and the right chart is for the iC3b 43 kDa bands. The PBS trial is the blue line, red line is for V62 against R53H and the green line is V62 against R78G.

Both plots from the previous experiment showed a dip in iC3b production at the 1:1 ratio of wild-type FH 1-4: mutant FH 1-4. But since these were not part of an overall trend, these results were regarded with suspicion (loading error, staining time *etc.* could also explain the dip). Nonetheless, this was investigated further as it may reflect the physiological situation in the plasma of a heterozygote for the mutation. In

the next study, C3b and FI (as in the previous experiment) were incubated with 40 nM, 30 nM, 20 nM and 10 nM total concentrations of FH 1-4 consisting of 0:1 or 1:0 (controls) and 1:1 mixtures of wild-type (V62) and mutant. In this case, the assay time was extended to four hours to increase the strength of the bands and thereby allow for more accurate density measurements.

Inspection of the resultant SDS-PAGE (Figure 3.17) and the densitometry readings (Figure 3.18) indicated that over the 24-fold extended assay period, there was detectable iC3b produced for both mutants (no wild-type FH 1-4 present) at 10 nM. Nonetheless, the mutants (alone) were, as expected, much less active than the V62 (wild type) FH 1-4. At 30 nM and 40 nM, the mixed samples (1:1 R53H:V62 and 1:1 R78G:V62) were equal in cofactor activity to the pure V62 samples. At 10 nM (total) the 1:1 mixtures exhibited less cofactor activity than pure V62 FH 1-4 but this is unremarkable since the mixtures contain only 5 nM V62 FH 1-4 and these concentrations of cofactor lie in the linear range (versus activity) according to the previous experiment (0-15 nM). The 20 nM mixtures exhibited an almost identical activity to 10 nM pure V62, in line with expectations. Therefore the dips in the plots shown on Figure 3.15 are likely to be anomalous and of no significance.

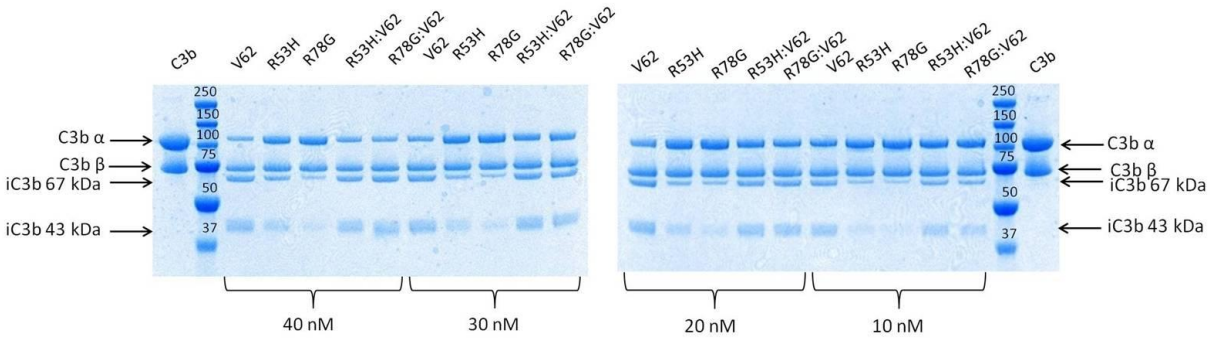


Figure 3.17 Homozygous vs. Heterozygous assay. The left SDS-PAGE shows the results for the 40 nM and 30 nM concentrations and the right shows 20 nM and 10 nM.

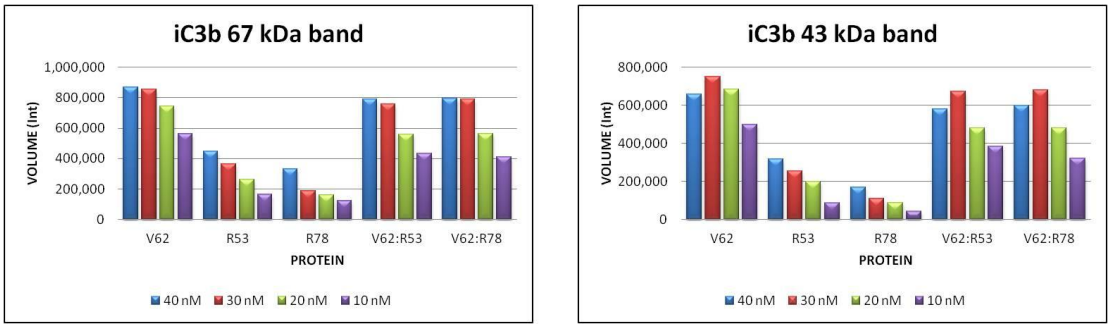


Figure 3.18 Quantification results for Homozygous vs. Heterozygous assay. The bar chart on the left shows the average result for the iC3b 67 kDa bands and the right hand chart the average result for the iC3b 43 kDa bands.

The fluorescence-based assay was also used to investigate differences in cofactor activity between pure and mixed samples at a total concentration of 4 µM FH 1-4. Thus the "heterozygous samples" contained 2 µM of V62 FH 1-4 plus 2 µM mutant. Concentrations of C3b and FI were 1.5 µM and 50 nM, respectively, in a final volume of 50 µl that also contained 100 µM ANS.

The outcome of these experiments (Figure 3.19) was similar to that of the SDS-PAGE assay (Figure 3.16). Unlike in the fluorescence assays shown in Figure 3.14, R78G is less active than R53H. But more importantly, the mixed samples showed no indication that the mutant FH 1-4 interferes with the wild-type FH 1-4. Taken together these results and combining with the SDS-PAGE results it seems that in these assays there is no destructive interference between mutant and wild-type versions of FH 1-4.

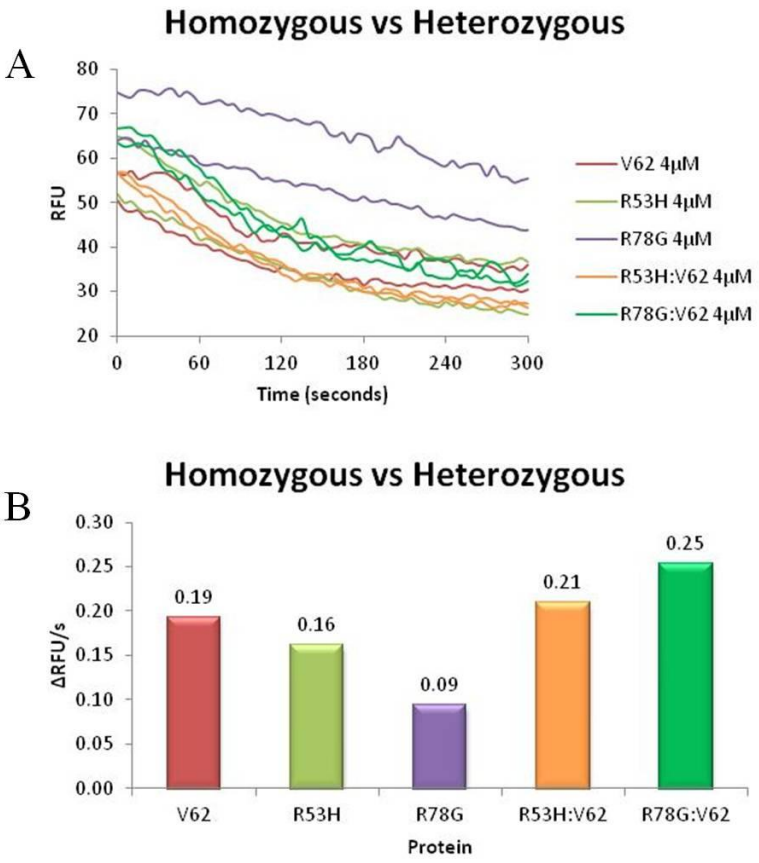


Figure 3.19 Fluorescence cofactor assay of Homozygous vs. Heterozygous. (A) Decrease in fluorescence as C3b becomes iC3b; (B) comparison of the initial rates of iC3b production.

3.3 Discussion

Our work showed that recombinant R53H FH 1-4 and R78G FH 1-4 (produced in *P. pastoris*), corresponding to aHUS-linked mutations in the N-terminal CCP of FH, have decreased cofactor activity and hence ability to regulate C3b amplification (and thereby control the AP) compared to FH 1-4 with the wild-type (V62) sequence, in the fluid phase. The ability of the mutants to bind C3b was investigated [53] and it was shown that R78G binds immobilized C3b poorly (K_D > 35 µM) whilst R53H can bind with a similar affinity as that seen for WT FH 1-4 (R53H K_D was ~ 12 µM compared to WT K_D of 10 µM)

The rare mutant R78G FH has been linked to aHUS [93] even though it occurs alongside wild-type FH in heterozygous patients. It is very likely that aHUS-linked R53H FH will also occur in heterozygous individuals, although this has not been

confirmed [93]. This suggested a dominant-negative effect in which the normal unaffected protein is prevented from working to some extent by the presence of the mutated protein [163, 164]. The work in this chapter was performed to investigate this suggestion using fluid-phase cofactor assays.

Two fluid-phase cofactor assays were used in the study, an end-time SDS-PAGE-based assay and a continuous fluorescence-based one. The SDS-PAGE assay is more commonly used [53, 140] as the results are easy to interpret visually in a qualitative sense. In trying to use this assay in a quantitative fashion during this study, a few challenges became apparent like the importance of minimizing any differences in gel staining and de-staining times and variations between batches of stain that were more difficult to control. The fluorescence assay was adapted from the original protocol [152] to use significantly smaller reaction volumes. Although the assay allowed the conversion of C3b to iC3b to be monitored in real time it too presented technical challenges. Using small volumes increased the risk of errors during pipetting. Controlling the temperature of the reaction was also an issue as the plastic 96-well plates was prone to cooling very quickly when removed from the plate reader for the purposes of adding solutions to the wells; to ameliorate this problem, all the proteins and the buffer were incubated at 37 °C. But the biggest challenge for the fluorescence assay was timing; The FI was added last but then the reaction mixture required agitation. During this mixing step (5 seconds) FI begins to act so that initial rate is hard to measure.

A series of incubations were set up for the SDS-PAGE based cofactor assay, with FH, FI and C3b, where the total concentration of FH was kept constant but the ratio of wild-type protein (V62 FH 1-4) to mutant protein (R53H FH 1-4 or R78G FH 1-4) was incremented over the range 1:0 to 0:1. This would have revealed any inhibitory effects of the mutant on the cofactor activity of the wild-type (V62) FH 1-4 since R53H and R78G are both very significantly poorer as cofactors than the wild-type material. The presence of the mutant proteins in the reaction mixture had no significant effects on the activity of V62 FH 1-4 in these assays. Similarly, in a fluorescence-based assay that affords the opportunity to observe C3b cleavage in real time, the mutants appeared to have no effect on the activity of the V62 FH 1-4.

In order to appreciate why both R53H and R78G are poor co-factors, the position of the mutations within the three-dimensional structure of CCP 1 must be considered (Figure 3.20 (A) and (B)). Arg 78 of R78G is within the binding interface for C3b. Substitution of the Arg side chain, which is polar and charged, to the non-polar neutral side chain of Gly could certainly explain the lower binding to C3b seen previously and the associated loss of cofactor activity [53]. Interestingly, in their subsequent crystal structure of the FH 1-4:C3b complex, Wu *et al*. reported the presence of a salt bridge between the guanidinium group of R78 and D732 of the α'NT domain in C3b (Figure 3.20 (C)) [57].

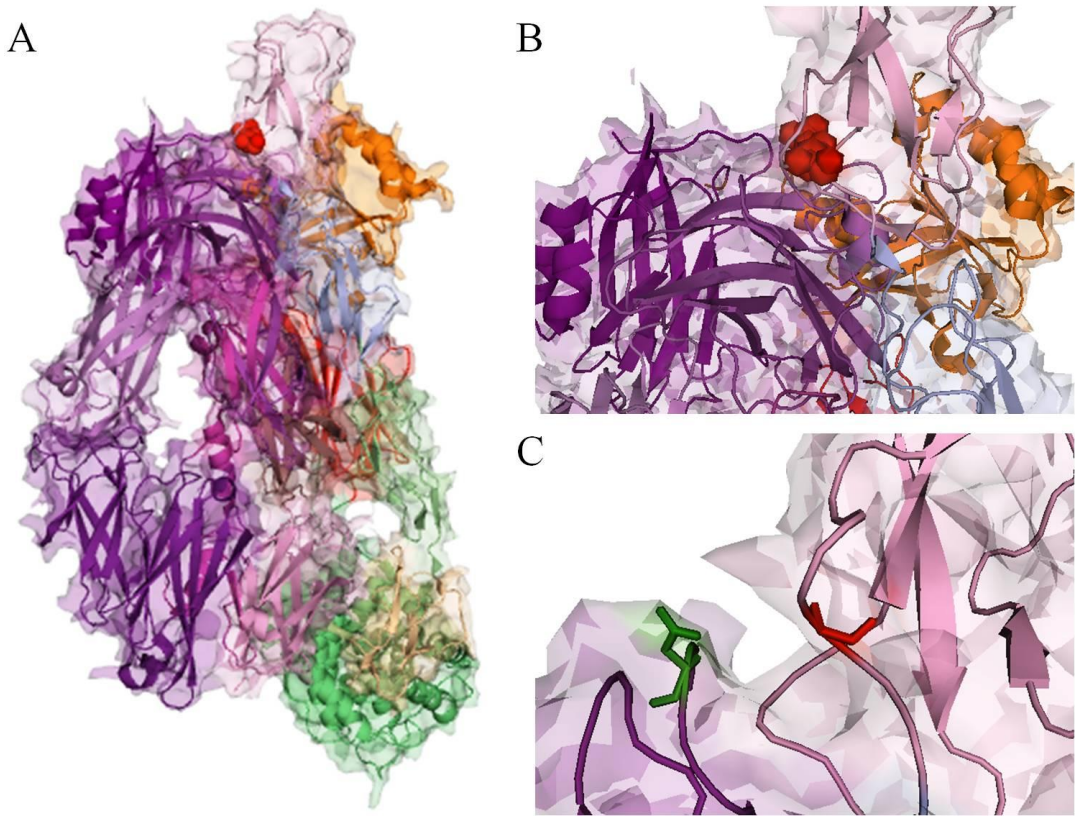


Figure 3.20 Location of R78G within the C3b-FH 1-4 complex structure solved by crystallography. (A) Complex of C3b and FH 1-4 (PDB_ID 2WII) with the position of R78G indicated by red spheres. (B) A zoomed-in view of the region of the complex in the vicinity of R78. (C) Salt bridge between FH 1-4 G78 (red sticks) and C3b D732 (green sticks). Images visualized with PyMOL.

In the case of R53H FH, the mutated residue is located on the opposite side of CCP 1, *i.e.* away from the C3b-binding interface. Petchl *et al* hypothesized that this could explain why R53H FH 1-4 has a similar C3b-binding affinity to V62 FH 1-4 [53].

Hence affinity for C3b is not the source of this mutant's lowered cofactor activity. Recently, Roversi *et al* [155] proposed a model of the C3b:FH 1-4:FI ternary complex based on the FI crystal structure [155]. This implied (Figure 3.21) a possible interaction between Arg53 and the FIMAC domain from the heavy chain of FI. Hence these findings suggest that R53H has a decreased cofactor activity due to the inability of the C3b:FH 1-4 (R53H) complex to bind to FI.

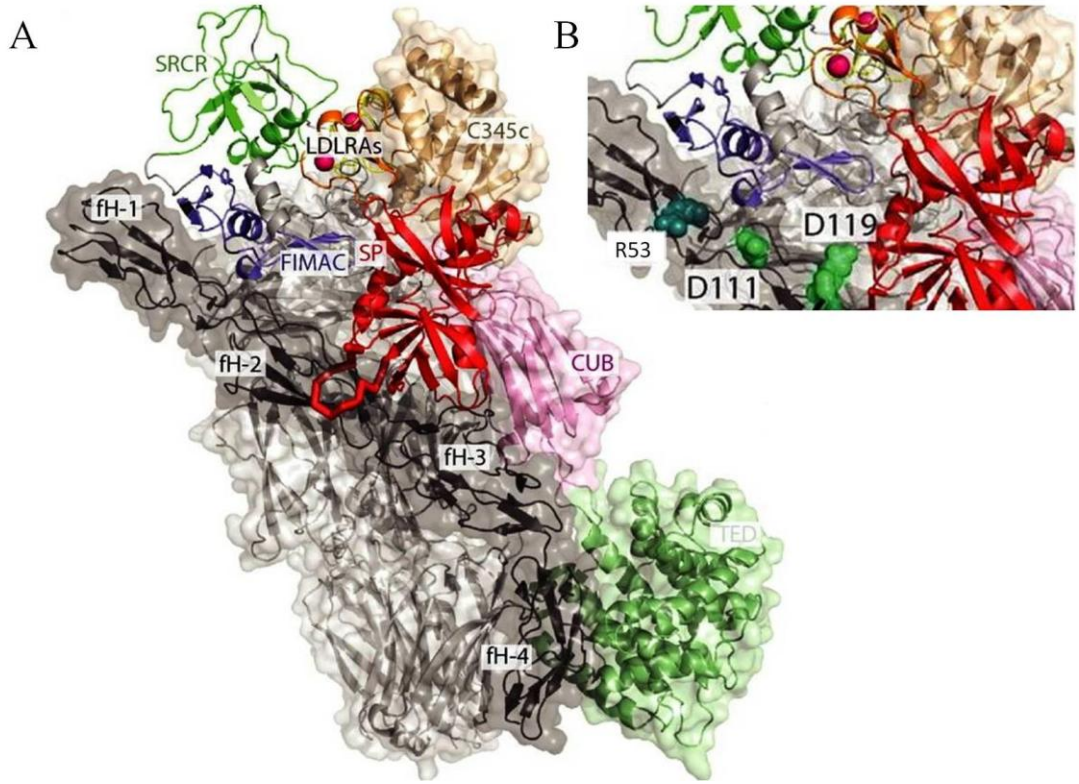


Figure 3.21 Location of R53H on the ternary complex of C3b, FH 1-4 and FI. In (A) is the model of the C3b-FH-FI complex with FI shown as a cartoon and C3b and FH shown as cartoons with semitransparent surfaces. (B) shows a zoomed in view of the interaction between FI and R53H (cyan). Figure (A) from [155] and (B) adapted from [155].

Taken together, all these structural and functional observations suggested that R53H FH (but not R78G) should destructively interfere with the function of wild-type FH since the two proteins will compete for C3b binding but only the complex of wild-type FH and C3b will be a substrate for FI. According to this hypothesis, R53H FH 1-4 should inhibit the cofactor activity of V62 FH 1-4, but this is not what we observed. This is despite the fact that the concentration of C3b:CFH 1-4 (wt) complex should (by design) have been limiting, since C3b was present in excess and

we were working at FH 1-4 concentrations that were below the K_d (10 μM) for C3b:FH 1-4, and were in a linear range with regard to the rate of iC3b production. Nonetheless, we have not fully disproved the "negative dominant hypothesis" due to limitations of the current study. First, these studies were carried out in fluid phase whereas the key actions of FH occur on surfaces bearing specific molecular markers that activate FH. Second, we were using just the N-terminal region of FH that does not include the C3b/C3d-binding and surface-binding sites in CCPs 7 and CCPs 19-20.

Chapter 4 Study of intramolecular distances within FH using EPR

The most powerful techniques for determining the structures of proteins are X-ray diffraction and NMR spectroscopy but neither of these is ideally suited to the study of FH. Although crystal structures have been published for portions or segments of FH [48, 57, 62], no successful attempts at crystallizing FH either alone or in the presence of C3b have been reported to date. With NMR the challenge is the high molecular weight and extended shape of FH; currently, due to broad lines and spectral complexity, pairs or triplets of CCPs [52, 55] represent the upper limit for structure determination.

Recently a very promising integrative approach has been developed to combine high-resolution structures derived from crystals and NMR with low-resolution techniques such as EM and SAXS [52, 165]. This allows the smaller high-resolution models to be fitted into shape envelopes generated form larger fragments of the protein. Further structural restraints – such as from chemical cross-linking, EPR or FRET – can also be included to improve the reliability and accuracy of the modeling process [59, 166]. In the current work, the use of EPR-derived distance restraints as part of an integrative structure determination of FH was investigated.

Early efforts were focused on the central region of FH since this is hypothesized (see section 1.6) to be critical for promoting a conformation of FH in which N-terminal and C-terminal C3b-binding sites are brought into proximity. The double module FH 12-13 was used initially for a "proof of concept" study since its NMR-derived structure had previously been determined, and SAXS data collected. The plan was to incorporate a paramagnetic group into each module and compare the EPR-derived distance measurement to the one predicted from the solution structure. It was anticipated that a major challenge to be overcome would be the one of site-specific labeling. Subsequently, if successful, this approach would be applied to longer fragments such as FH 10-13, FH 10-15 and FH 11-14 for which combined NMR/SAXS-based models are already available. Ultimately it would be applied to full-length FH.

4.1 Preparation of the DNA required for site-directed spin labeling of recombinant FH fragments.

With the need for attachment of paramagnetic groups in mind, mutagenesis (see Methods, section 2.1.1.4) was used to place either an additional cysteine residue or a transglutaminase tag (PKPQQFM) at either the C terminus or the N terminus of each recombinant FH fragment or segment.

A note on the naming convention for FH fragments used hereafter: the first letter after the FH module numbers within the recombinant segment corresponds to the addition of a residue or residues at the N terminus and the second letter to the C terminal addition; *e.g.* FH 11-14 CT has an additional cysteine at its N terminus and a transglutaminase tag inserted at its C terminus, while TT or CC would indicate transglutaminase tags or cysteine residues, respectively, at both termini. WT signifies "wild type" proteins or, strictly speaking, native-sequence fragments. The mutated DNA constructs were ligated into the desired *E. coli* vector (see Methods, section 2.1.4) and sequenced prior to transformation into *P. pastoris* (see Methods, section 2.1.1.3).

4.2 Production and purification of proteins

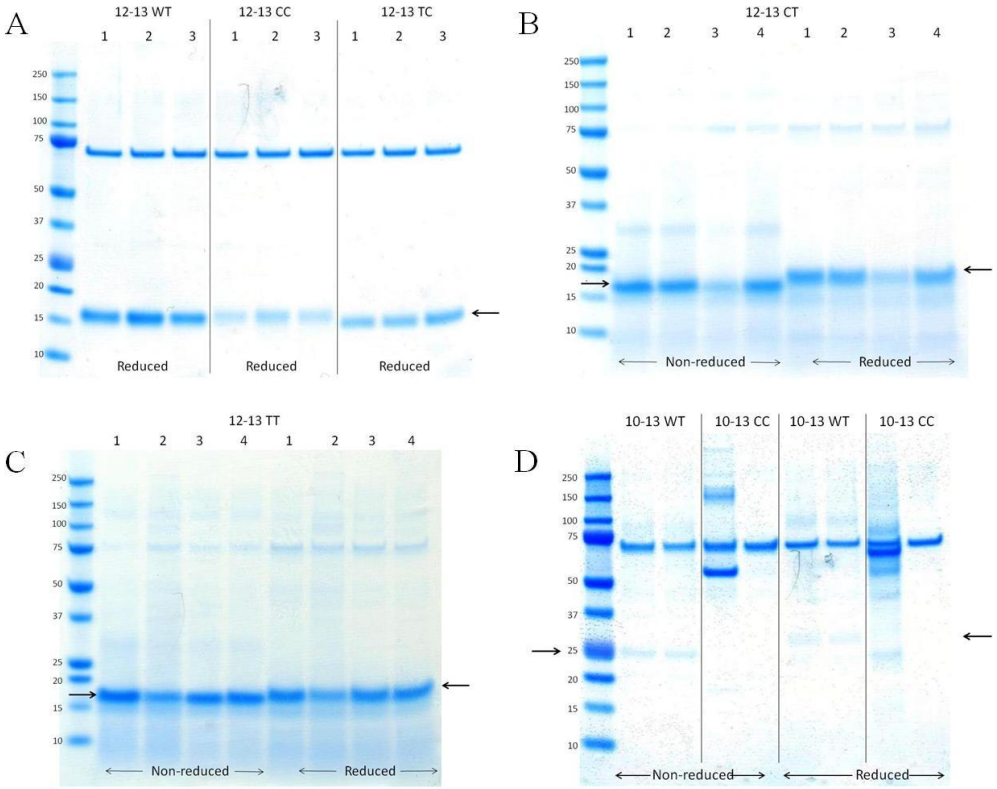
4.2.1 Small-scale test expressions.

Twelve DNA constructs (see Table 4.1) coding for the modified FH segments were transformed successfully into *P. pastoris* (see Methods section 2.2.3) and cells grown on agar plates containing Zeocin. For each construct, between two and four colonies were subsequently screened for protein production by inoculating 100 ml cultures (see Methods section 2.2.4) that were grown up in a shaking incubator. Protein production was assessed by SDS-PAGE (Figure 4.1) after enzymatic cleavage of N-glycans with $EndoH_f$ (MW = ~75 kD).

Table 4.1 Details of modified FH fragments for EPR.

FH fragment	No of CCPs	N-terminus tag	C-terminus tag
12-13 WT	2	-	-
12-13 CC	2	Cys	Cys
12-13 TT	2	TTag	TTag
12-13 CT	2	Cys	TTag
12-13 TC	2	TTag	Cys
10-13 WT	4	-	-
10-13 CC	4	Cys	Cys
10-13 TT	4	TTag	TTag
10-13 CT	4	Cys	TTag
10-13 TC	4	TTag	Cys
11-14 CT	4	Cys	TTag
10-15 CT	6	Cys	TTag

Coding used: Cys = cysteine and TTag = transglutaminase tag.



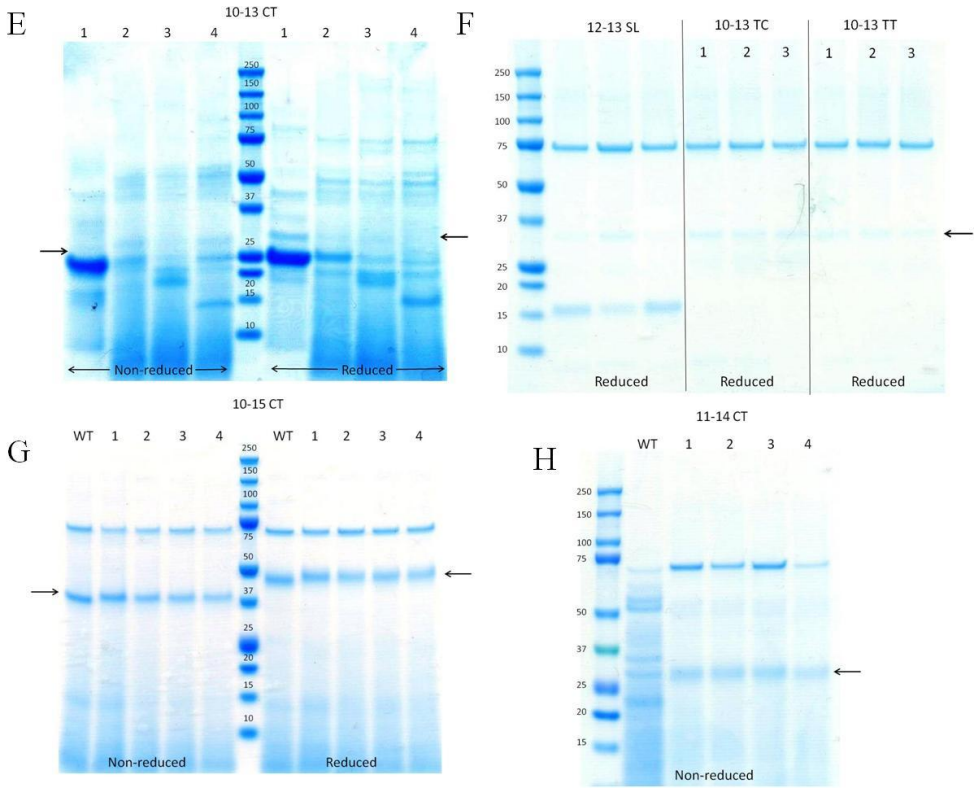


Figure 4.1 Test expressions for all EPR constructs. SDS-PAGE was used to assess protein production by two to four colonies of each protein in 100 ml cultures. The bands visable at 75 kDa, in all gels, is Endo H_f while the arrows indicate the bands of interest. 12-13 SL is discussed in Chapter 5.

Test expressions for a set of FH 12-13 constructs showed production, by all the colonies, of a protein that migrated at the expected positions of ~15 kDa under reducing conditions.

On the other hand, mixed results were obtained for the four-CCP FH 10-13 constructs. Candidate bands for native-sequence FH 10-13 (WT) were detected on polyacrylamide gels following electrophoresis under reducing and non-reducing conditions. But only one of the FH 10-13 CC colonies appeared, on this basis, to produce protein and yield was extremely poor. The FH 10-13 CT colonies also failed to produce any candidates that migrate as would be expected for a protein of the expected size during SDS-PAGE. To further investigate these colonies, the growth medium (after spinning out cells) from colony 1 was fractionated using size-exclusion chromatography for analysis by mass spectrometry. This suggested that the main protein product of FH 10-13CT was between 5 and 6 kDa smaller than

expected. In the cases of both FH 10-13 TC and FH 10-13 TT, protein production was achieved but at low levels compared to FH 10-13 WT.

In the cases of FH 10-15 CT and FH 11-14 CT all four colonies produced protein of the expected size.

As none of the FH 10-13 proteins designed for labeling had been produced successfully they were not subjected to any further investigation. Likewise, with regard to the FH 12-13 proteins, only FH 12-13 CC and FH 12-13 TT were taken forward for further study. Table 4.2 summarizes the results of the test expressions.

Table 4.2 Summary of results from test productions.

Construct	Estimated MWt (Da)	Outcome
11-14 CT	29000	Good yield, used for large scale protein production
10-15CT	42000	Good yield, used for large scale protein production
10-13 WT	28000	Low yield, not used in further experiments
10-13 CC	28000	Poor yield, not used in further experiments
10-13 TT	29000	Poor yield, not used in further experiments
10-13 CT	29000	MWt incorrect; to be investigated further
10-13 TC	29000	Poor yield, not used in further experiments
12-13 WT	14000	Good yield, but not needed for further experiments
12-13 CC	14000	Good yield used for large scale protein production
12-13 TT	15000	Good yield used for large scale protein production
12-13 CT	15000	Good yield, but not used for further experiments
12-13 TC	15000	Good yield, but not used for further experiments

Molecular weights (MWt) listed in Table 4.2 have been estimated based on the calculated result from Protparam. The reason for this is for unexplained reasons the

proteins (with the exception of FH 12-13 CC) failed to ionize during the recording of mass spectrometry experiments.

4.2.2 Production and purification of FH 12-13 CC

Production of FH 12-13 CC was carried out in shaker flasks (see Methods section 2.2.5) with an initial volume of 4 L of BMGY media (500 ml per 2 L flask) which was reduced to 1 L of BMMY media (125 ml per 2 L flask). After removal of cells by centrifugation, the supernatant was adjusted to pH 4.0 and diluted fivefold prior to application to a 30 ml SP Sepharose chromatography column (containing cation-exchange resin, see Methods section 2.3.1). The bound protein was then eluted from the column using a 1 M NaCl gradient, and 2 ml fractions were collected and analyzed by SDS-PAGE (Figure 4.2) based on 20 μL aliquots (each was treated with Endo H_F prior to preparation for loading on to the gel).

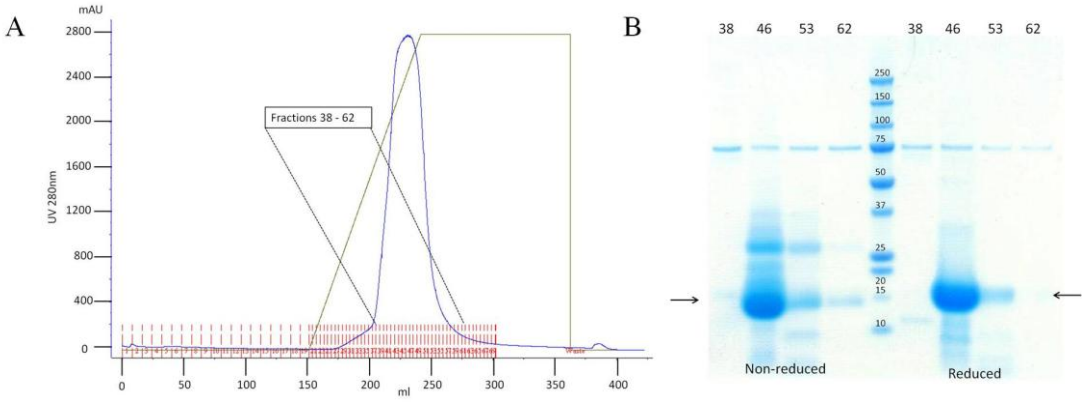


Figure 4.2 Purification of FH 12-13 CC on SP Sepharose column. (A) The elution profile (blue line) using a 1 M NaCl gradient (green line). (B) SDS-PAGE for the fractions run under both non-reducing and reducing conditions. Arrows indicate the FH 12-13 CC bands; the band visible at 75 kDa is Endo H_f.

A single broad peak with a very high OD reading was eluted between 200 and 275 ml and SDS-PAGE confirmed the protein was present within fractions 46 to 53. Under reducing conditions a strong single band was present at the expected size for monomeric protein, but suspected dimer (26-kDa band) was present under non-reducing conditions. Fractions 42-53 were pooled (giving a final volume of ~ 50 ml), then a 20 ml aliquot of this pool was concentrated to 2.0 ml, and buffer-exchanged into PBS for application to a size-exclusion column. The remaining 30 ml was stored

at -80 °C. Two injections onto the size-exclusion chromatography column of 1 ml were performed.

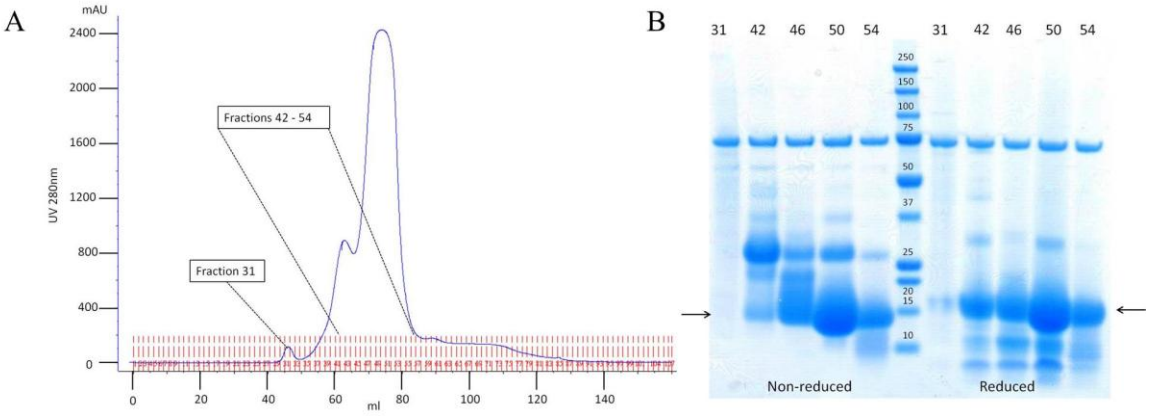


Figure 4.3 Size-exclusion chromatography of FH 12-13 CC using HiLoad 16/60 Superdex 75, first of two 1.0 ml applications. (A) Elution profile for injection; (B) the gradient SDS-PAGE with the lane numbers corresponding to the fractions indicated on (A). The-75 kDa band is Endo H_f.

Following each of the 1 ml injections, two poorly resolved peaks were present in the chromatogram, eluting between about 60 ml and 75 ml (see Figure 4.3 (A), only the result of the first injection is shown). Fraction 42 contains a high proportion of a putative dimer of FH 12-13 CC (that becomes monomeric under reducing conditions). Fractions 46-54 covered the 75 ml peak and reflected the incomplete resolution of monomer and dimer, while there was also evidence of "clipping" *i.e.* proteolytic cleavage products that become evident under reducing conditions. These have probably become more prominent over the course of the purification procedure due to the continuing action of proteases in the stored incompletely purified samples. Finally, fractions 46-54 from both SEC injections were pooled (final volume of ~ 20 ml) and then treated with Endo H_F before concentration to 1.0 ml, and reapplication to the size-exclusion chromatography column.

The elution profile following this second pass through a size-exclusion column was dominated by three poorly resolved peaks, with the main one at 75 ml, as before (Figure 4.4 (A)). Analysis of the fractions by SDS-PAGE (Figure 4.4 (B) and (C)) suggested that the first peak contains trimeric FH 12-13 CC, the second peak contains a dimeric form and the third peak contained monomeric FH 12-13 CC (under both reducing and non-reducing conditions) although the "clipped" fragments

remain evident under reducing conditions. On this evidence, fractions 50-56 were pooled (fraction 49 was excluded due to a small amount of dimer being present) and concentrated to 1.0 ml, yielding a concentration of 200 µM (2.8 mg/ml), before being stored at – 20 °C. From 1 L of BMMY the approximate yield is 50 mg/L.

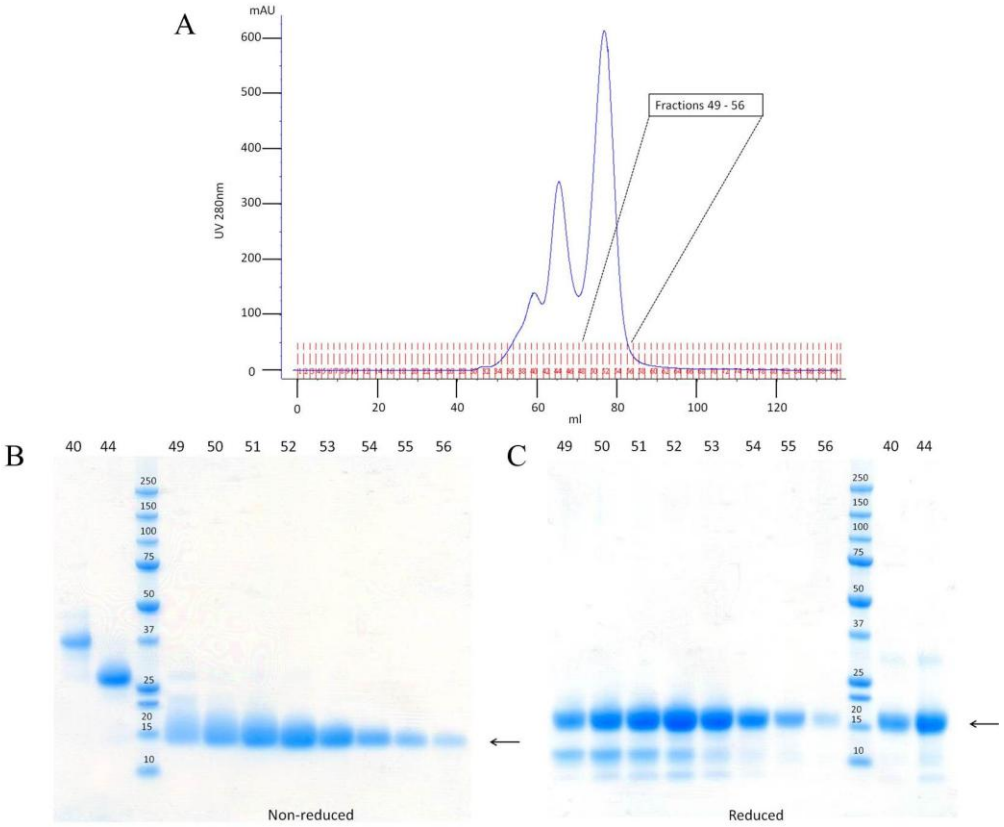


Figure 4.4 Size-exclusion chromatography of FH 12-13 CC using HiLoad 16/60 Superdex 75, second application. (A) The elution profile; (B) The gradient SDS-PAGE for the fractions indicated on (A) under non-reducing conditions; (C) Fractions subjected to SDS-PAGE under reducing conditions. Arrows indicate the bands of interest.

4.2.3 Purification of FH 12-13 TT

Expression and initial purification of FH 12-13 TT, and analysis of the results by SDS-PAGE, was carried out in an identical manner to that of FH 12-13 CC.

The elution profile form the SP-Sepharose column showed a very large and broad peak with some leading-edge shoulders (Figure 4.5). Analysis of the peaks by SDS-PAGE (15 µl loaded) showed that fractions 12, 18 (first shoulder) and 24 (second shoulder) contained no protein. Fractions 30-42, corresponding to the main peak,

contained large amounts of protein at the expected sizes although under reducing conditions small amounts of clipped protein are visible.

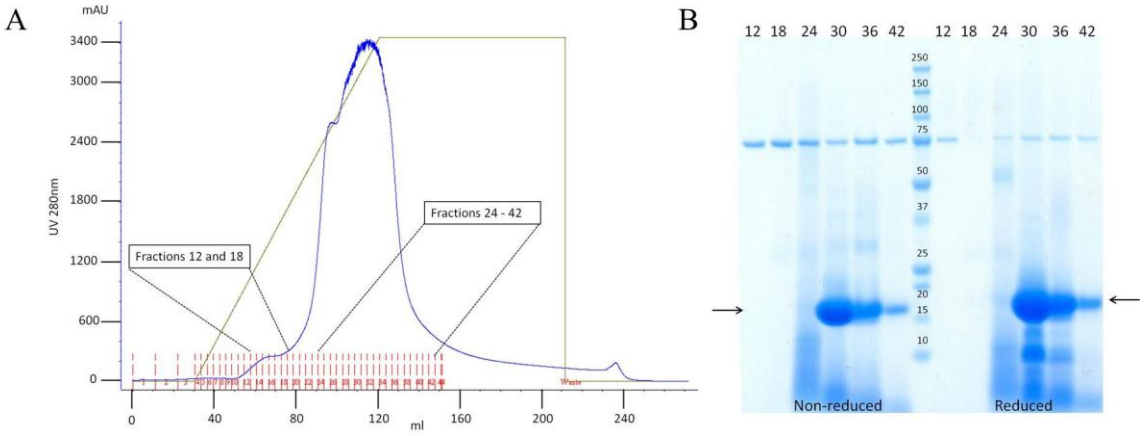


Figure 4.5 Ion-exchange chromatography of FH 12-13 TT on SP Sepharose. (A) The elution profile (blue line) using a 1 M NaCl gradient (green line). (B) Gradient SDS-PAGE for indicated fractions run under both non-reducing and reducing conditions. The putative FH 12-13 TT bands are indicated by arrows. The band visible at 75 kDa is Endo H_f.

Fractions 28-42 were pooled (~ 60 ml) and then half was stored whilst the remaining half was treated with Endo H_f before being concentrated to 2 ml for size-exclusion chromatography; four injections of 0.5 ml were performed.

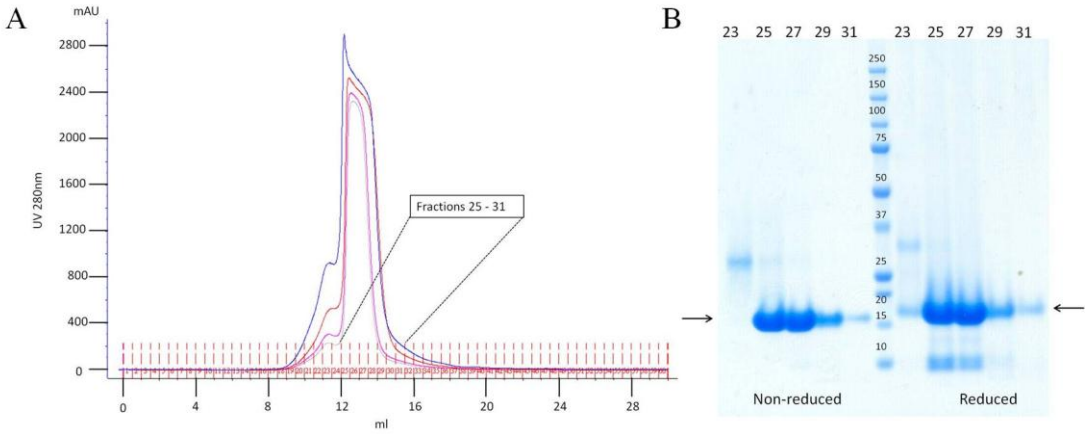


Figure 4.6 Size-exclusion chromatography of FH 12-13 TT on Superdex 75 HR 10/30. (A) The four elution profiles are shown overlaid; (B) gradient SDS-PAGE for the first injection. The putative FH 12-13 TT bands are indicated by arrows.

For all four injections a dominant peak was obtained with a leading-edge shoulder (Figure 4.6 (A)). The SDS-PAGE (Figure 4.6 (B)) for the shoulder (fraction 23)

showed protein that could correspond to FH 12-13 TT dimer although under reducing conditions there were two bands visible, the lower one probably being monomeric protein. Fractions 25-31, corresponding to the main peak, contained protein at the expected size with evidence of "clipped" material under reducing conditions. These fractions (25-31) were combined and concentrated to 300 µM (4.5 mg/ml) in 1 ml and stored at -20 °C prior to being labeled. From 1 L of BMMY the approximate yield is 4.5 mg/L.

4.2.4 Purification of FH 11-14 CT

Protein production was, again, carried out in shaker flasks with an initial volume of 2 L (4x 500 ml) of BMGY media, which was then reduced to 500 ml of BMMY media. As before the pH of the harvested supernatant was adjusted to 4.0 and diluted fivefold before being applied to a 30 ml SP Sepharose column.

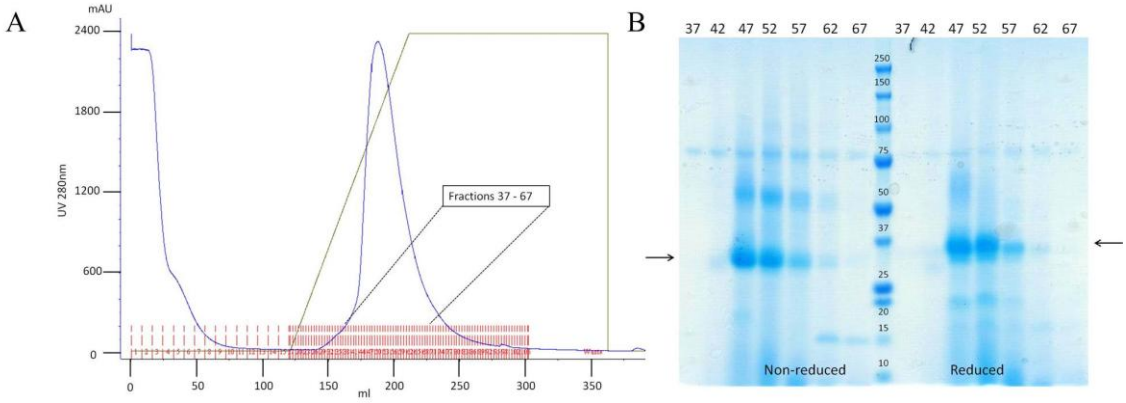


Figure 4.7 Ion-exchange chromatography of FH 11-14 CT on SP Sepharose. (A) Elution profile (blue line) using a 1 M NaCl gradient (green line). (B) Gradient SDS-PAGE of fractions (indicated in (A)) under both non-reducing and reducing conditions. The protein bands are indicated by arrows and the 75-kDa band visible is Endo H_f.

The elution profile (Figure 4.7 (A)) showed a single peak eluting between 500 and 900 mM NaCl. According to SDS-PAGE (Figure 4.7 (B)) fractions 47-62 contained protein of the appropriate size and there were combined for Endo H_f treatment then concentrated to 1.0 ml for size-exclusion chromatography using two 0.5 ml injections.

Four peaks appeared in the elution profile from size-exclusion chromatography (Figure 4.8 (A), injection one only); the first peak contained no protein whilst the

second peak contained protein at the expected size for a dimer of FH 11-14 CT under non-reducing conditions (gel not shown). The third, most prominent, peak contained protein of the expected size for monomeric FH 11-14 CT (Figure 4.8 (B)). Under non-reducing conditions the protein gave a single band but under reducing conditions there was evidence that clipped versions of the protein were also present.

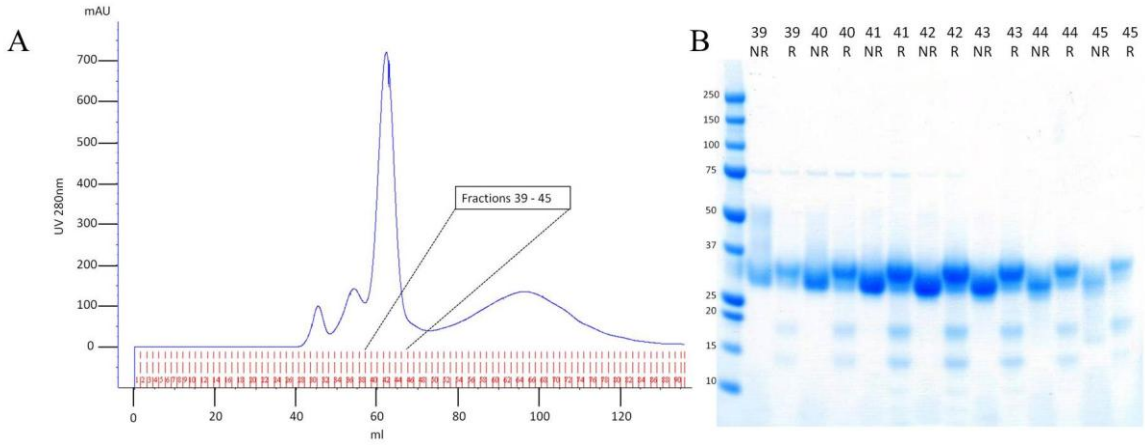


Figure 4.8 Size-exclusion chromatography of 11-14 CT using HiLoad 16/60 Superdex 75. (A) The elution profile. (B) The fractions indicated in (A) were then run under non-reducing (NR) and reducing (R) conditions on gradient SDS-PAGE.

In an attempt to remove the 75-kDa impurity from the sample both sets of fractions 39-45 were combined and concentrated to 1.0 ml for re-application to the size-exclusion column. The result was a broad peak with shoulders (Figure 4.9 (A)). Analysis of the fractions on SDS-PAGE (Figure 4.9 (B)) showed fractions 40 and 41 still contained a small amount of the higher-molecular weight impurity so these were discarded; fractions 42-48 were combined and concentrated to a final volume of 1.0 ml (concentration was 330 µM, 9.6 mg/ml) and then stored at -20 °C. From 0.5 L of BMMY the approximate yield is 10 mg/L.

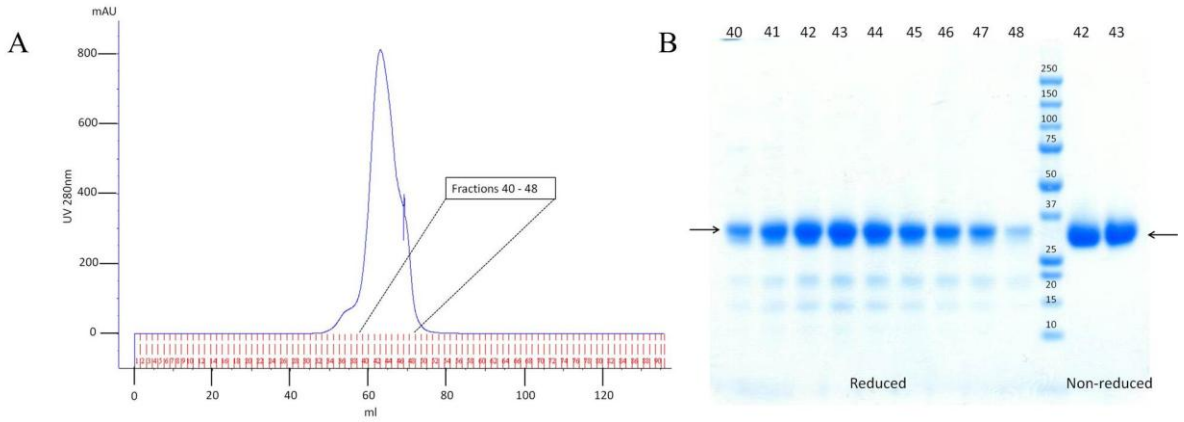


Figure 4.9 FH 11-14 CT: Second size-exclusion chromatography step performed on HiLoad 16/60 Superdex 75. (A) Elution profile. (B) The fractions indicated in (A) were then run in reducing conditions on SDS-PAGE alongside fraction 42 and 43 run under non-reducing conditions. Arrows indicate the expected position on the gel of the target protein.

4.2.5 Purification of FH 10-15 CT

The production of FH 10-15 CT was carried in a similar manner to that of FH 12-13 CC. The resulting supernatant was concentrated to 200 ml of which 20 ml was applied to HiLoad 16/60 Superdex 75 size-exclusion column (each injection contained 1 ml). Previous work (by Dr Christoph Schmidt, [167]) showed that FH 10-15 WT eluted from an identical column at 60 ml so the peak eluting at 60 ml (Figure 4.10 (A)) was investigated by SDS-PAGE. A small amount of each fraction from the peak was treated with Endo H_f prior to being run on the gel (Figure 4.10 (B)), which confirmed that the eluted protein migrated as would be expected for FH 10-15 CT (~40 kDa) although double bands were obtained (also seen by Schmidt). Pooled fractions (39-43) were combined for all 20 injections and (volume ~ 110 ml) treated with Endo H_f.

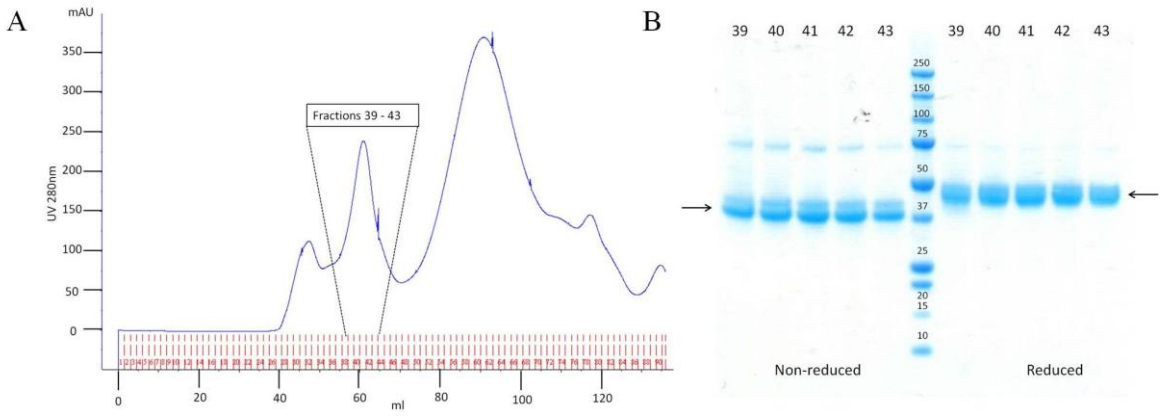


Figure 4.10 First purification step for FH 10-15 CT using HiLoad 16/60 Superdex 75. (A) The elution profile for injection one. (B) Gradient SDS-PAGE analysis of the fractions indicated in (A). The putative FH 10-15 CT band is indicated by arrows; the 75-kDa band is Endo H_f.

After deglycosylation the protein was concentrated to a final volume of 2.0 ml and two 1.0 ml injections were reapplied to the size-exclusion column. One major peak with a shoulder was obtained in the chromatogram (Figure 4.11 (A)). Analysis of fraction 36, which corresponded to the shoulder, showed it contained what is presumably dimeric protein under non-reducing conditions that became monomeric when reduced. Analysis of the remaining fractions (Figure 4.11 (B)) showed fraction 38 contained both monomer and dimer protein under non-reducing conditions but fractions 40 and 42 contained only monomeric protein. Under reducing conditions fractions 40-46 contained a small amount of clipped protein. Nonetheless, fractions 40-46 from both runs were combined and concentrated to a final concentration of 200 µM (8.4 mg/ml) in 1 ml prior to being stored at - 20°C. From 1 L of BMMY the approximate yield is 8 mg/L.

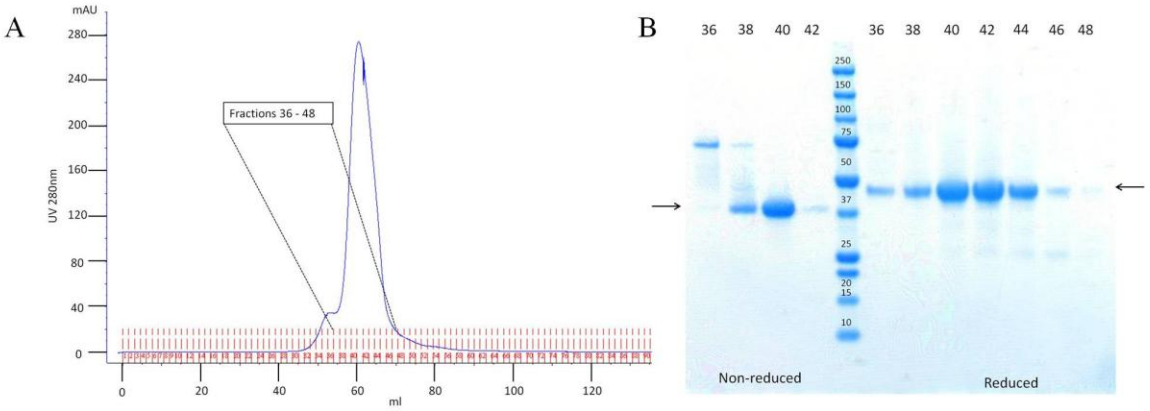


Figure 4.11 Second purification of FH 10-15 CT using HiLoad 16/60 Superdex 75. The size-exclusion chromatogram is shown in (A). For SDS-PAGE in (B) the lane numbers indicate the fractions numbers (also indicated in (A)); arrows indicate the expected migration positions for FH 10-15 CT under non-reducing and reducing conditions.

The yields obtained for FH 12-13 CC, FH 12-13 TT, 11-14 CT and FH 10-15 CT are summarized in Table 4.3

Table 4.3 Summary of final concentrations obtained from the large-scale production.

FH construct	Final concentration (µM)
12-13 CC[A]	200
12-13 TT[B]	300
11-14 CT[C]	330
10-15 CT[D]	200

All concentrations are in a final volume of 1 ml in PBS buffer. [A, B] and [D] from 1 L of BMMY (50 mg/L, 4.5 mg/L and 8 mg/L respectively). [C] from 0.5 L of BMMY (10 mg/L)

4.3 Distance measurements using DEER

Labeling of FH 12-13 TT was carried out using a sample of 100 µl of protein at 100 µM and 20x excess of the transglutaminase spin label overnight at room temperature. Excess spin label was removed using size exclusion chromatography on a Superdex 75 10/30 column.

As the "proof-of-concept" protein, the FH 12-13 TT sample (100 µl of 50 µM in 10 mM Tris-HCl dissolved in 99.9 % D_2O, pH 7.4 plus 50% deuterated glycerol) was studied in initial DEER experiments using a Bruker E580 EPR machine running

pulsed X-band EPR. Analysis of the data was performed by Dr David Norman, University of Dundee using DeerAnalysis2011 [168]. The original data is shown in Figure 4.12.

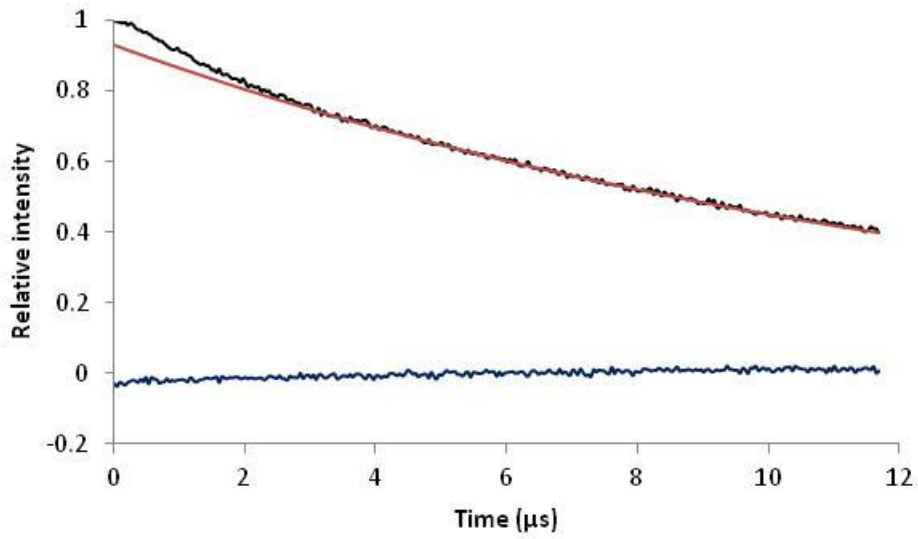


Figure 4.12 Initial DEER experiment showing raw data for FH 12-13 TT. The black line represents FH 12-13 TT with the red line being the fit. Background is shown by the dark blue line.

After background removal and applying Tikhonvo regularization to the data [169] (Figure 4.13), the process of extracting distances measurements could begin. Analysis of the data, however, highlighted a difficulty with the sample. The drop in the relative intensity of oscillation, from 1.0 to 0.92, is much smaller than the expected drop, which would typically be from 1.0 to around 0.6 (personal communication with Dr David Norman, University of Dundee). The main reason behind for this was thought to be that the protein is not efficiently double labeled.

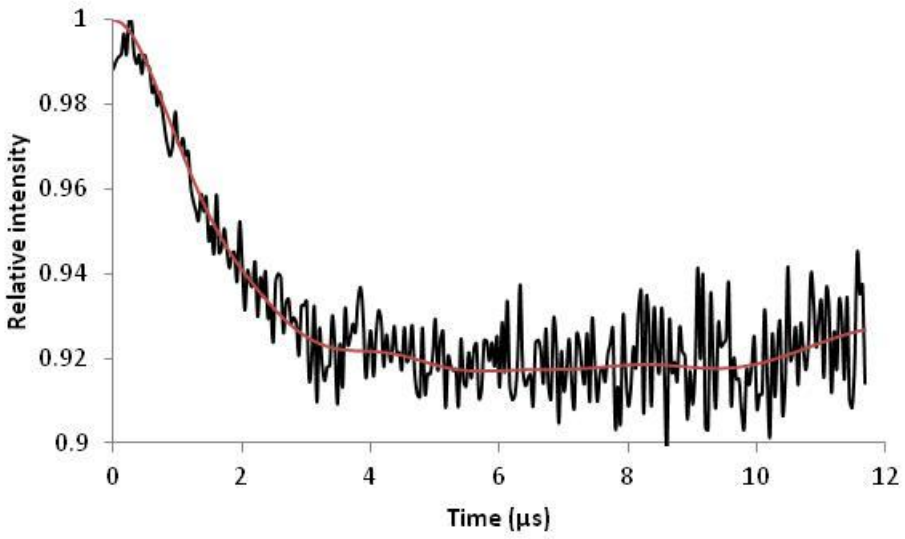


Figure 4.13 Background-corrected trace for FH 12-13 TT and Tikhonov fitting. Corrected time trace is shown in black and the Tikhonov fitting in red.

Nonetheless, it was possible to extract inter-label distances from the data (Figure 4.13). In the NMR-derived structure of FH 12-13 (PDB: 2KMS) the N-terminal and C-terminal residues - Thr_{690} and Ser_{804} - are ~ 5.2 nm apart (Figure 4.14). Allowing for the peptide tags to which the spin labels are attached, suggests a total distance in the vicinity of ~ 6 nm (measured between Gln_{686} and Gln_{808}).

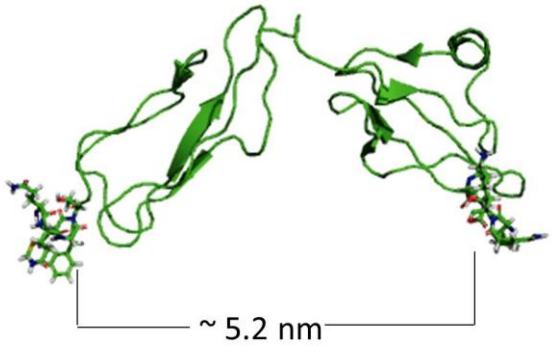


Figure 4.14 Model of FH 12-13 with transglutaminase tags. The FH 12-13 model (PBD: 2KMS) is visualized as a cartoon and the transglutaminase tags as sticks. The distance was measured using PyMOL.

This distance estimation does not take into account the flexibility of the tag sequence or of the spin label itself. A distance-distribution plot (Figure 4.15) was created for FH 12-13 TT after Tikhonov regularization. Multiple peaks were obtained

suggesting a possible distance of between 4.8 and 7 nm. An accurate distance cannot therefore be predicted, presumably due to the dynamic nature of the sample as well as the suspected low percentage of doubly labeled protein. Nonetheless, the DEER-derived distances are compatible with the distance between N-terminal and C-terminal residues measured in the NMR-derived structure.

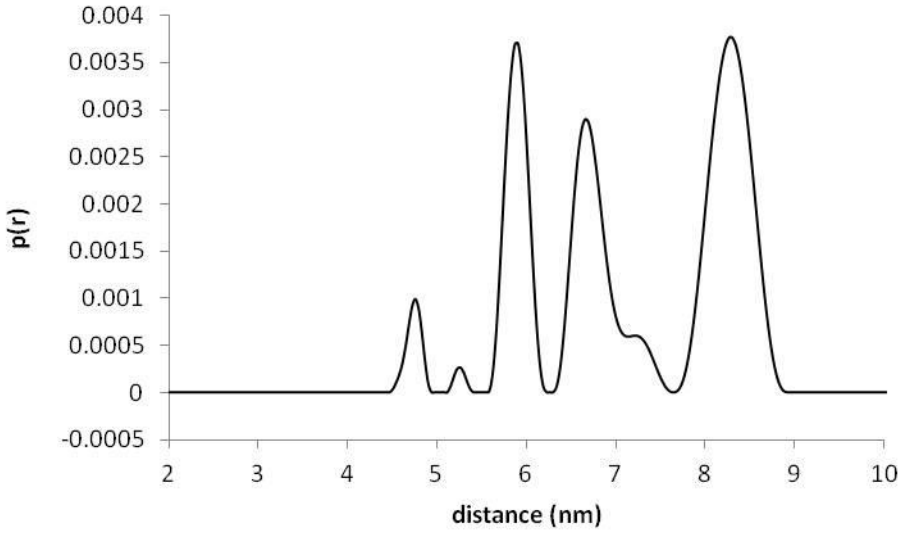


Figure 4.15 FH 12-13 TT DEER-derived distance distribution. The plot shows the range of possible distances obtained for FH 12-13 TT. The multiple peaks probably arise from the low percentage of doubly labeled protein as well as from the dynamic nature of the sample.

For both FH 11-14 CT and FH 10-15 CT, the proteins (100 µM) were firstly labeled with 20x excess of MTSSL at room temperature for 4 hours and then dialyzed to remove the unused label. The protein was then labeled with the transglutaminase label (as for FH 12-13 TT) and removal of the excess label was performed using size exclusion.

The DEER experiments performed on the labeled samples of FH 11-14 CT and FH 10-15 CT were, unfortunately, unsuccessful. Figure 4.16 shows the original data for FH 11-14 CT and Figure 4.17 for FH 10-15 CT. Neither trace showed any drop in intensity. The most likely reason for this is that the proteins were not in fact doubly labeled. Another possible explanation was the low concentration of the proteins used to collect data.

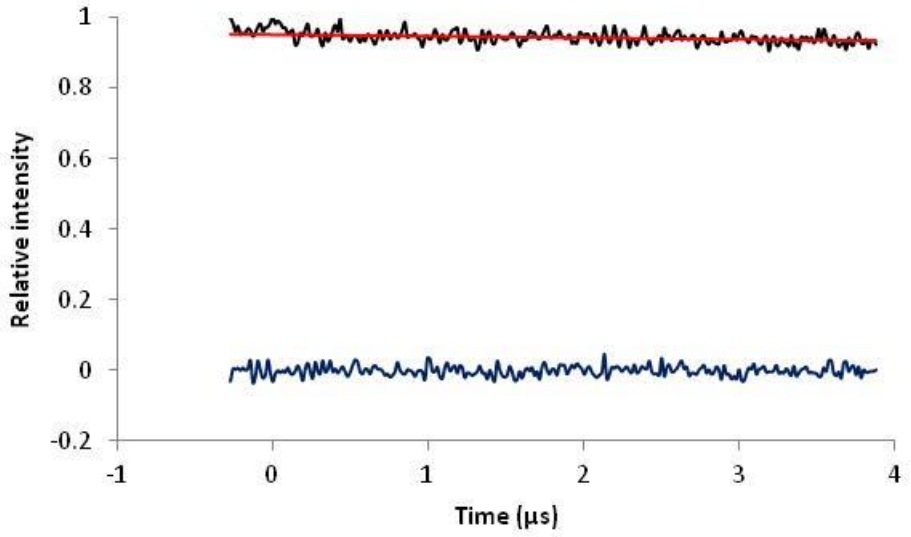


Figure 4.16 Raw DEER data for FH 11-14 CT. Represented by the black line is FH 11-14 CT with the red line being the fit. Background is shown by the dark blue line.

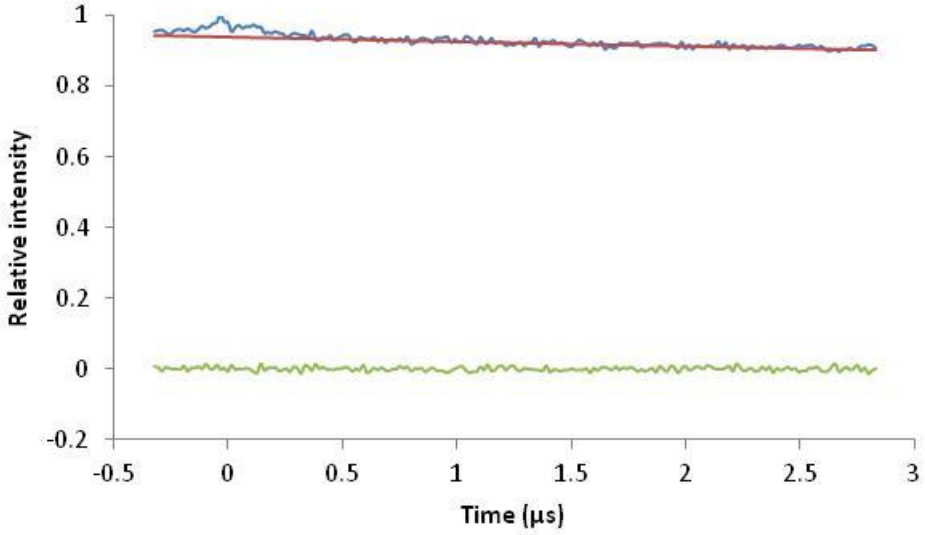


Figure 4.17 Raw DEER data for FH 10-15 CT. Blue line represents FH 10-15 CT with the red line being the fit. Background is shown by the green line.

A further attempt to gather distance measurements was initiated using FH 12-13 CC. A sample containing 100 µM of FH 12-13 CC was incubated with a 20x excess of the MTSSL spin label for four hours at room temperature. The results of labeling of FH 12-13 CC after removal of the excess label was assessed using mass spectrometry (Figure 4.18). This showed that two species had originally been present

(before labeling), differing by a GlcNAc groups (200 Da). Only one of the cysteines had been successfully labeled in each of the two species. The simulated mass of two labels and either one or two GlcNac groups did not correspond to any of the majors peaks in the mass spectrum of the labeled protein. The labeled protein sample was submitted for DEER analysis but no distance measurement was achieved, consistent with the absence of a doubly-labeled sample. In order to confirm that both sulfhydral groups were available for labeling another sample was prepared but this time labeled with maleimide. The mass spectrometry data (Figure 4.19) showed that the majority of the protein remained unlabeled with a small amount showing the presence of a single label. A peak was obtained for double labeled protein but it is very small and could even be considered noise. Unfortunately labeling of both cysteines was never achieved despite further attempts (increase in incubation time as well as amount of reduction agent present) so no further work was continued with this protein

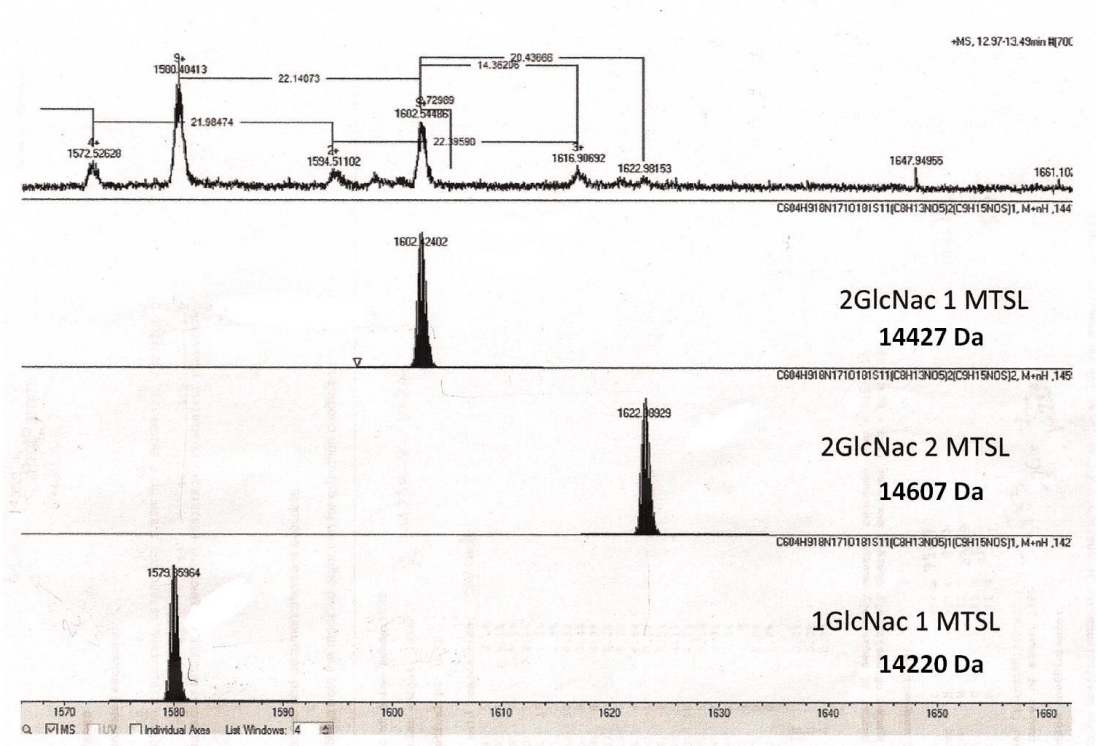


Figure 4.18 LC-MS spectrum for FH 12-13 CC to check MTSSL labeling efficiency. The raw data is shown above the simulated masses corresponding to the different labeling states achieved.

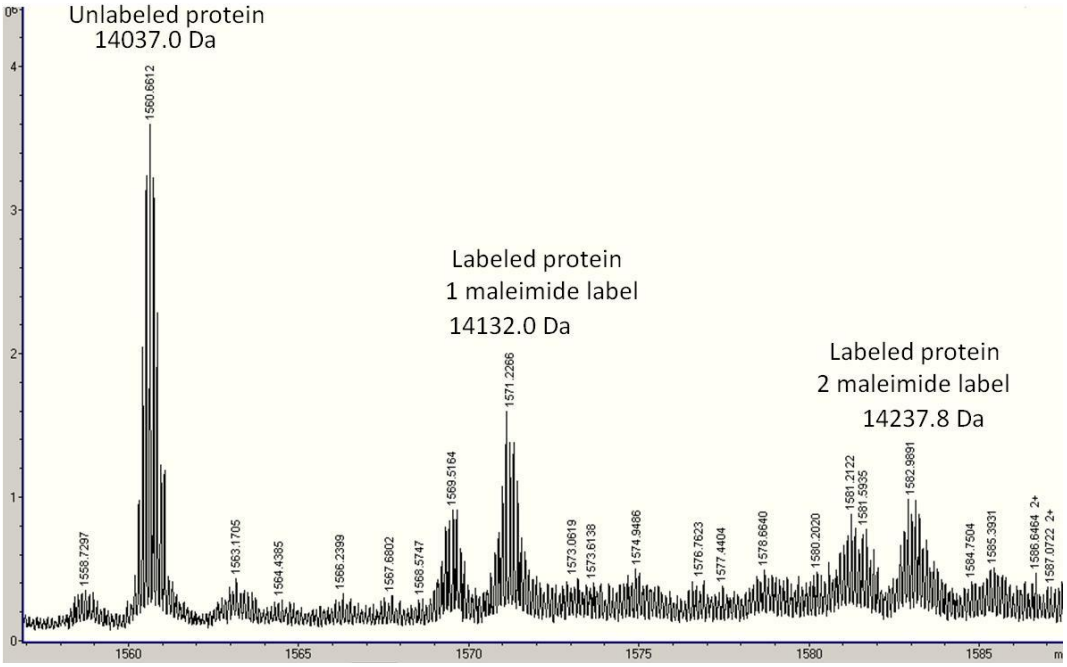


Figure 4.19 FH 12-13 CC labeled with maleimide. The spectrum shows the results of the use of maleimide to confirm if the sulfhydral groups are available for labeling.

4.4 Conclusions

The aim of the work presented in this chapter was to prove the principle that it will be possible to obtain EPR-derived distances measurements [121, 139] which will help to define the three-dimensional structure of the large complement regulatory protein, FH. A major challenge to this approach is the need for site-specific attachment of paramagnetic species. This is normally achieved [170, 171] through the insertion of cysteine residues using mutagenesis. While this is straightforward for proteins that do not contain disulfides bridges [172] it is more problematic in the case of a protein like FH that has 80 cysteines and 40 disulfides. This is because of the potential for disulfide scrambling as a result of the presence of extra non-native cysteines. Moreover, once the mutated protein is correctly folded the free cysteines can cause oligomerisation meaning that a careful selection of labeling conditions is required. As an alternative to the engineered-in cysteine route that avoids some of these potential pitfalls, the use of transglutaminase has its own problems since it involves the use of flexible N-terminal or C-terminal peptides and hence uncertainty in the precise position of the paramagnetic label with respect to the rest of the

protein. Given these challenges, it was decided to instigate EPR-based studies of FH by attempting to collect data on small recombinant segments from its central region that are known to form compact structures [55].

Twelve constructs were prepared which allowed for a combination of labeling techniques to be explored. The results for the various FH 12-13 constructs were very promising as all exhibited good levels of production during initial trials. To date, only FH 12-13 CC and FH 12-13 TT have been used for labeling experiments as it was decided to prioritize the proteins that could be labeled in a single step; there was insufficient time to attempt labeling of FH 12-13 CT and FH 12-13 TC but these samples have been stored for future experiments. The FH 12-13 CC results therefore demonstrated, for the first time, incorporation of cysteines (at both termini), without apparently disrupting disulphide bonds and this provides hope that such a feat could be repeated for larger segments and even intact FH in due course.

For FH 12-13 TT, it proved possible to extract DEER-derived distances and compare them to what would be expected based on the NMR solution structure of FH 12-13 WT. The result, however, should be treated with a degree of caution as the labeling of the sample was not very efficient. To try and improve this it is likely that more transglutaminase enzyme was required during the labeling process so any future experiments would involve finding the right amount of enzyme for each protein to achieve the highest percentage of labeling possible.

Of the FH 10-13 constructs, only FH 10-13 WT produced a useful quantity of protein during trial growths. The reasons for lack of non-WT FH 10-13 production were not investigated due to lack of time and due to the availability of alternative constructs (shorter and longer) for proof-of-principle work. One possibility is that there is an upper limit on the number of cysteines that can be tolerated for this approach. Since none of the FH 10-13 constructs created for labeling were successful during the trials they were removed from the study.

Of the two longer constructs, the production of both FH 11-14 CT and FH 10-15 CT was successful, but unfortunately neither was able to give distance measurements.

Possible reasons are, sample concentrations were too low or that the majority of the sample was only singly labeled but further investigation is required.

EPR is now becoming a popular tool to aid structure determination of proteins [139] but is still relatively unused for the study of complement proteins with only DAF being studied so far [166]. After correcting all the issues identified by the current study, future work would move on to labeling larger fragments (e.g. FH 8-15) and then trying to measure various distance in the central region in order to gather more restraints for NMR and SAXS modeling.

A method not used in the study but one which may aid future work is the insertion of unnatural (i.e. non-encoded) amino acids within the protein sequence. The insertion of the unnatural amino acid has two advantages over the techniques used in the study; first it removes the risk of disrupting the natural disulphides within the CCPs and second the flexibility within the transglutaminase tag is not an issue, reducing the errors in the distance measurements.

An orthogonal labeling strategy might involve exploitation of the N-glycans present as post-transcriptional modifications in both recombinant and plasma-derived FH. These are mainly - but not exclusively - located in the central modules of FH and display sugars rings that contain diols. The diols could be used for tagging with a spin label that has been appropriately derivatised with a hydrazine group [173] [174]. Prior treatment with Endo H_f would result in a single GlcNAc attached to each N-glycan site; the number and position of these could be varied by protein engineering to achieve double labeling at strategic points in the protein.

Chapter 5 Investigating the architectural role of the central region of FH

The segment of FH encompassing CCPs 8-18 appears to have no major binding sites for either C3b or GAGs [63]. The lack of binding sites in this region suggests that it plays a structural, or architectural, role in positioning the binding sites at either terminus so that they can cooperate in engaging ligands. It is noteworthy that intermodular linking sequences (linkers) within the CCPs 10-15 region are generally longer (5-8 residues) and CCPs generally smaller (e.g. CCP 13 is the smallest CCP at 51 amino acid residues) compared with N-terminal and C-terminal regions of FH. This could be consistent with a hinge-like role for these central CCPs. NMR-derived solution structures are available for double-module constructs consisting of CCPs 10-11, 11-12 and 12-13 [52, 55]. All three adopt "V-shapes" (Figure 5.1). For example, FH 12-13 [55], which contains the smallest CCP as well as the longest linker (eight amino acid residues), exhibits an ~80° tilt between the modules (measured using a convention in which 0° would correspond to a fully extended conformation of the two modules). Six out of eight residues in the linker contain bulky side chains and four of these are buried in the interface between the two modules that "glues" them together and impedes motion between the modules.

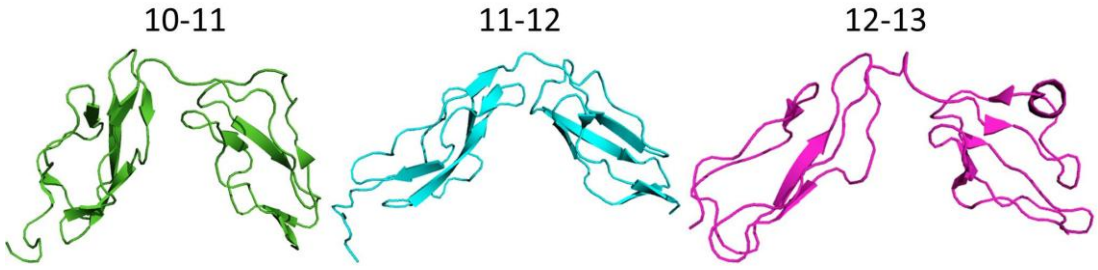


Figure 5.1 NMR-derived solution structures of module pairs from central region of FH. PDB models used: FH 10-11 (4B2R), FH 11-12 (4B2S) and FH 12-13 (2KMS).

In the current study, a mutagenesis investigation was designed to assess the functional and structural roles of the long linker between CCPs 12 and 13 of FH. Initial work was aimed at studying the effects of linker mutations in the context of FH 12-13 since this would facilitate a detailed comparison by NMR of the structure

and dynamics of mutant and wild-type proteins [55]. Subsequently, the same mutations were introduced into FH 10-15 and FH 8-15 to allow assessment of the effects of a mutated linker, using low-resolution structural methods [52, 55], on longer segments of FH. The final goal was to mutate the CCPs 12-13 linker within full-length (20-CCP) FH.

5.1 Design of the new linkers for FH 12-13

The original linker between FH 12-13 contains eight residues: Val_{745} – Ala_{746} – Ile_{747} – Asp_{748} – Lys_{749} – Leu_{750} – Lys_{751} – Lys_{752}. The residues Val_{745}, Ala_{746}, Ile_{747} and Leu_{750} all have hydrophobic side chains that are buried in the interface between the modules. In designing a linker that would be expected to be less rigid, the residues $Ile_{747,}$ Lys_{749} and Leu_{750} were each mutated to glycine. While retaining the length of the linker, this should disrupt the interface between the two modules and hence promote significant intermodular flexibility. The name assigned to this linker was "three-glycine" (3xGLY). Another mutant was created in which Ala_{746}, Ile_{747}, Asp_{748}, Lys_{749} and Leu_{750} were all deleted. It was reasoned that the remaining three-residue linker would disrupt the V-shape of FH 12-13 but would be too short to allow much flexion between modules. This linker was referred to as "short-linker" (SL).

5.2 Preparation of DNA encoding FH 12-13, FH 10-15, FH 8-15 and full-length FH

Polymerase chain reactions were successfully used to amplify DNA segments encoding native-sequence FH 12-13 (360 bp), FH 10-15 (1083 bp) and FH 8-15 (1446 bp) from the full-length FH-coding synthetic DNA gene (designed by Dr Christoph, prepared by GeneArt) which had been codon/expression optimized for *P. pastoris* [140].

The codon expression-optimized (for *P. pastoris*) DNA encoding full-length FH containing the short 12-13 linker (*i.e.* FH SL) was synthesised and cloned into pPICZαB by GeneArt (Life Technologies Ltd.). The DNA encoding FH 10-15 SL, FH 8-15 SL and FH 3xGLY was successfully generated in the current study from the codon-optimized full-length FH DNA from [140]using fusion PCR [175].

DNA encoding FH 12-13 SL was prepared using mutagenesis and cloned into pPICZαB by former lab member, Dr Mara Guariento.

The PCR products were subject to restriction digestions and gel purification prior to ligation into pPICZαB. Colonies from the ligation reactions underwent PCR-based screening to ensure insertion of the product had occurred (as opposed to re-ligation of the empty vector). Successful ligation should yield bands at ~550 bp for FH 12-13, ~1300 bp for FH 10-15, ~ 1600 bp for FH 8-15 and 3600 bp for full-length FH; empty pPICZαB should give a band of 330 bp. The results of both the amplification PCRs and screening PCRs are shown in Figure 5.2. Where bands of the expected size were obtained, the appropriate colonies were chosen for plasmid extraction and the resulting DNA was then sequenced.

The DNA containing the 3xGLY linker in FH 12-13, FH 10-15 and FH 8-15 was achieved using site-directed mutagenesis (see methods section 2.1.1.4). To verify that the mutagenesis was successful the resulting DNA was sequenced.

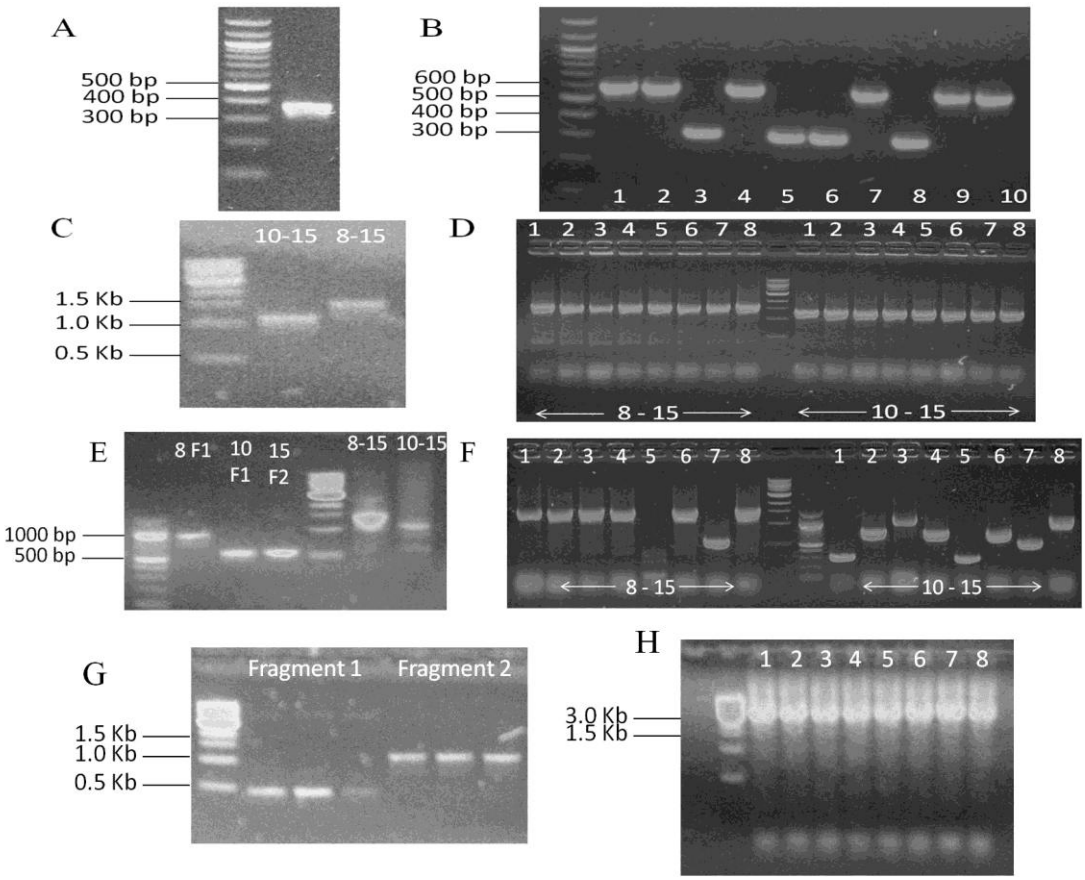


Figure 5.2 Amplification and screening PCR of various recombinant FH Segments. The agarose gels in (A) and (C) show the successful amplification of FH 12-13 WT (A, 360 bp), FH 10-15 WT (C, 1083 bp) and FH 8-15 WT (C, 1446 bp) from codon-optimized DNA encoding full-length native sequence FH. In (B) is the screening result for FH 12-13 with lanes 1, 2, 4, 7, 9 and 10 corresponding to pPICZαB containing the FH 12-13 insert. Lanes 3, 5, 6, and 8 all contain empty pPICZαB. The screening results for FH 10-15 WT and FH 8-15 WT are shown in gel (D); all eight colonies screened for both constructs showed successful incorporation of the inserts with pPICZαB. In (E) and (G) is shown the amplification of the fragments need for fusion PCR to generate 8-15 SL, 10-15 SL and FH 3xGLY. The screening results for 8-15 3xGLY and 10-15 3xGLY are shown in (F) with colonies 1-4 being correct (6-8 not sequenced) for 8-15 3xGLY and 2, 4 and 6 being positive for 10-15 3xGLY. For FH 3xGLY all colonies screened were positive (H).

A summary of how all the DNA constructs were prepared is shown in Table 5.1

Table 5.1 Production of DNA for FH constructs.

FH construct	Codon-optimized	DNA source	PCR	Site-directed mutagenesis
12-13 WT	YES	Codon optimized full length FH gene [2]	YES	-
12-13 3xGLY	YES	Codon optimized DNA encoding for 12-13 WT	-	YES
12-13 SL	NO	Non-codon optimized full length FH gene	YES[1]	-
10-15 WT	YES	Codon optimized full length FH gene [2]	YES	-
10-15 3xGLY	YES	Codon optimized DNA encoding for 10-15 WT	-	YES
10-15 SL	YES	Codon optimized DNA encoding for 10-15 WT	YES	-
8-15 WT	YES	Codon optimized full length FH gene [2]	YES	-
8-15 3xGLY	YES	Codon optimized DNA encoding for 8-15 WT	-	YES
8-15 SL	YES	Codon optimized DNA encoding for 8-15 WT	YES	-
FH WT	YES	Synthetic gene [2]	-	-
FH 3xGLY	YES	Codon optimized full length FH gene [2]	YES	-
FH SL	YES	Synthetic gene	-	-

[1] was prepared by Dr Mara Guariento and [2] by Dr Christoph Schmidt.

5.3 Expression and purification of proteins

The DNA for all of the wild type, 3xGLY linker and SL linker constructs was successfully transformed into the KM71H strain of *P. pastoris* (methods section 2.2.3). A small-scale protein production trial was carried out on between two and four colonies in the case of each construct (methods section 2.2.4). The resulting SDS-PAGE gels and western blots are shown in Figure 5.3.

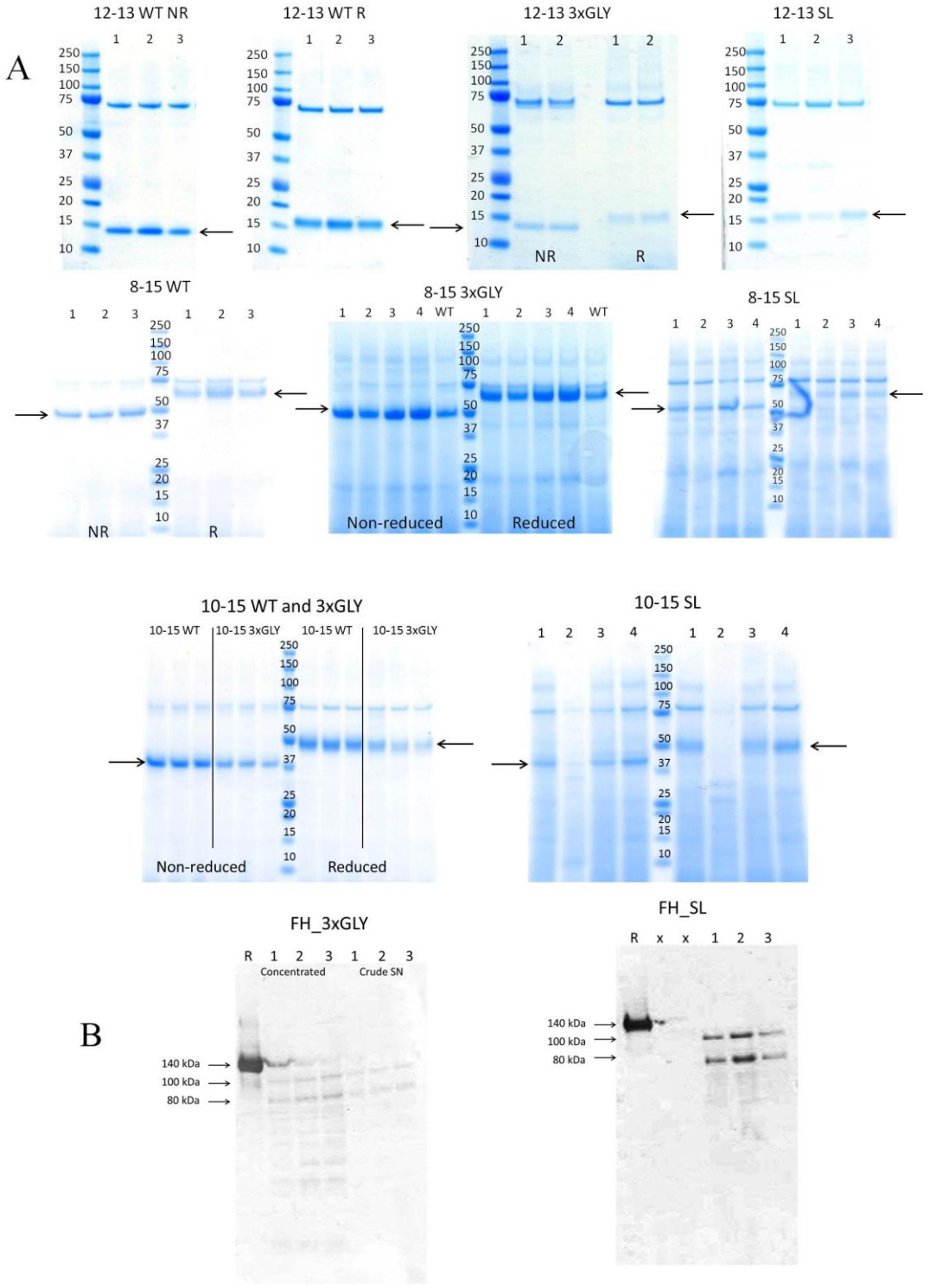


Figure 5.3 Small-scale protein production trials of all *P. pastoris* linker constructs. (A) Gradient SDS-PAGE; the band at 75 kDa detected in all gels is Endo H_f. Arrows indicate the bands of interest. NR means non-reducing, and R means reducing conditions. (B) Western blots; Lane "R" in each gel contains FH from Complement Technologies (used here as a positive control); in the FH 3xGLY gel, lanes 1-3 correspond to the colony number and are fivefold concentrated or are crude supernatant, as indicated. For the FH SL gel, lanes 1-3 correspond to colony numbers while lanes marked x are empty.

Prior to loading onto SDS-PAGE, all samples were treated with Endo H_f to deglycosylate them. The small-scale production attempts for all of the FH 12-13 constructs were promising. Note that, as expected, the reduced samples tend to run slightly higher than equivalent size markers. Similarly, colonies containing DNA for FH 8-15 wild type, FH 8-15 3xGLY and 8-15 SL all produced proteins that migrate according to expectations. The trial productions of both FH 10-15 wild type and FH 10-15 3xGLY were also successful. For 10-15 SL only colonies 1, 3 and 4 produced protein successfully. The western blots for the full-length FH mutants were only run under non-reducing conditions, so a band was expected at ~100 kDa. All the colonies for FH 3xGLY as well as FH SL produced some protein that ran on the gel at the desired size although there was also a lower MWt contaminant band present at 80 kDa.

5.3.1 ^{15}N Fermentation

In order to record HSQC spectra of FH 12-13 wild type[2], FH 12-13 3xGLY and FH 12-13 SL, the first step was to produce isotopically labeled protein samples. The most efficient way to achieve this was with one-litre fermentations where the media was enriched with ^{15}N-ammonium sulfate (see Methods section 2.2.6). Agitation rates, dissolved oxygen levels, pH and temperature were monitored over the time course for each of the fermentations. Figure 5.4 shows the logs obtained for FH 12-13 WT and FH 12-13 SL (data for FH 12-13 3xGLY not shown due to software issues) which were as expected based on previous FH fermentation. After the addition of the cells (position A) a rapid increase in agitation occurs in order to maintain the oxygen level at 40% as the cells begin to consume the initial glycerol. Once the cells have consumed all the available glycerol, additional glycerol feeds are performed (position B). Before induction to methanol the temperature was reduced to 15°C to decrease proteolysis, once the temperature was reached the methanol was added (position C) where the agitation increased which indicated the cells were consuming the methanol. Additional methanol feeds were performed (position D) over 72 hours until the cells were harvested (position E).

[2] ^{15}N fermentation and purification of FH 12-13 WT was carried out by Kate Fisher, MSc student under the supervision of the author.

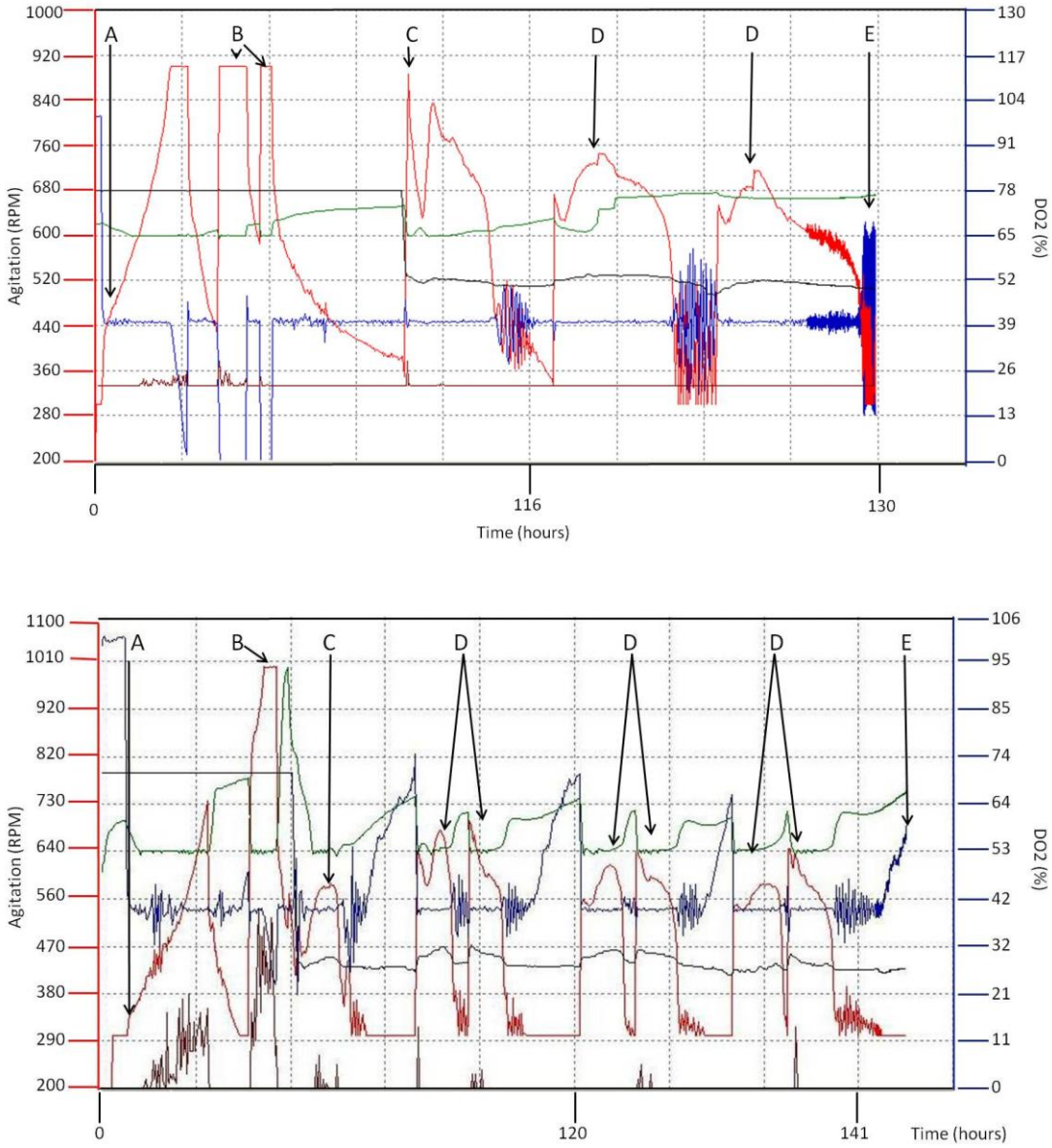


Figure 5.4 Log for fermentations of ^{15}N-labelled FH 12-13 wild type and FH 12-13 SL. The log for FH 12-13 wild type is shown above the log for FH 12-13 SL. In both are shown agitation (red), percentage of dissolved oxygen (DO_2, blue), pH (green), temperature (30°C -15°C after induction to methanol, black) and potassium hydroxide feed (used to keep pH at 5, brown). The (A) position indicates the addition of cells (B) is for glycerol and PTM1 salts feed, (C) is induction to methanol (D) indicates methanol and PTM1 salts feeds and (E) is when the cells were harvested.

5.3.2 Purification of FH 12-13 wild type

After harvesting, the supernatant was diluted threefold and adjusted to pH 4.0 before application to a 30 ml SP-Sepharose (cation-exchange chromatography) column. The protein was eluted using a 1 M NaCl gradient and the peaks were analyzed by SDS-PAGE (Figure 5.5). The chromatogram showed a major peak eluting at 700 mM

NaCl with a leading-edge shoulder. Analysis of the fractions by SDS-PAGE showed that fractions 51-78 from the main peak contained protein of the expected size. The higher band seen in fractions 69-78 could be dimer as it is not seen under reducing conditions; the lower bands seen in the reduced fractions are likely to be "clipped" species of FH 12-13 wild type. To conserve material, all of the fractions (numbers 54-78) containing the protein were recovered, combined and treated with Endo H_f.

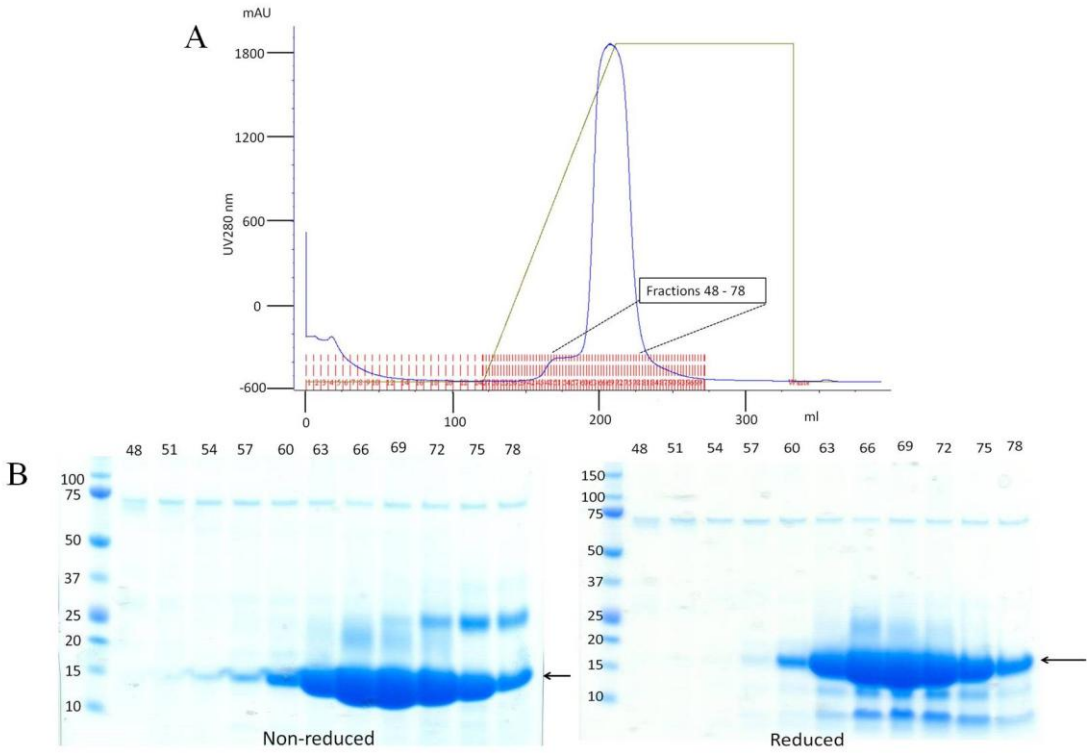


Figure 5.5 FH 12-13 wild type eluted from SP-Sepharose. (A) The elution profile (blue line) using a 1 M NaCl gradient (green line). (B) Gradient SDS-PAGE of fractions in both non-reducing and reducing conditions. Arrows indicate expected position of FH 12-13. The band visible at 75 kDa is Endo H_f.

After treatment with Endo H_f, the pooled fractions (50 ml) were concentrated to 1 ml then buffer exchanged into PBS to a final volume of 1 ml. The sample was then divided into two 0.5 ml samples for size-exclusion chromatography. Both of the size-exclusion profiles (Figure 5.6, (A)) show three peaks; a higher-molecular weight contaminant (ignored) followed by a small peak, then the main protein peak. The resulting SDS-PAGE gel from the first injection (Figure 5.6, (B)) confirmed that fractions 48-56 from the main peak contained FH 12-13. In the reduced samples the

lower-molecular weight clipped proteins were still present. The SDS-PAGE for injection 2 (Figure 5.6 (C)) gave the same result for the main peak fractions as injection 1. Samples containing fraction 44 were also run and showed that the smaller peak eluting ahead of the main peak was indeed dimer as the non-reduced sample had a size of ~ 30 kDa which when reduced ran at the expected size of 15 kDa.

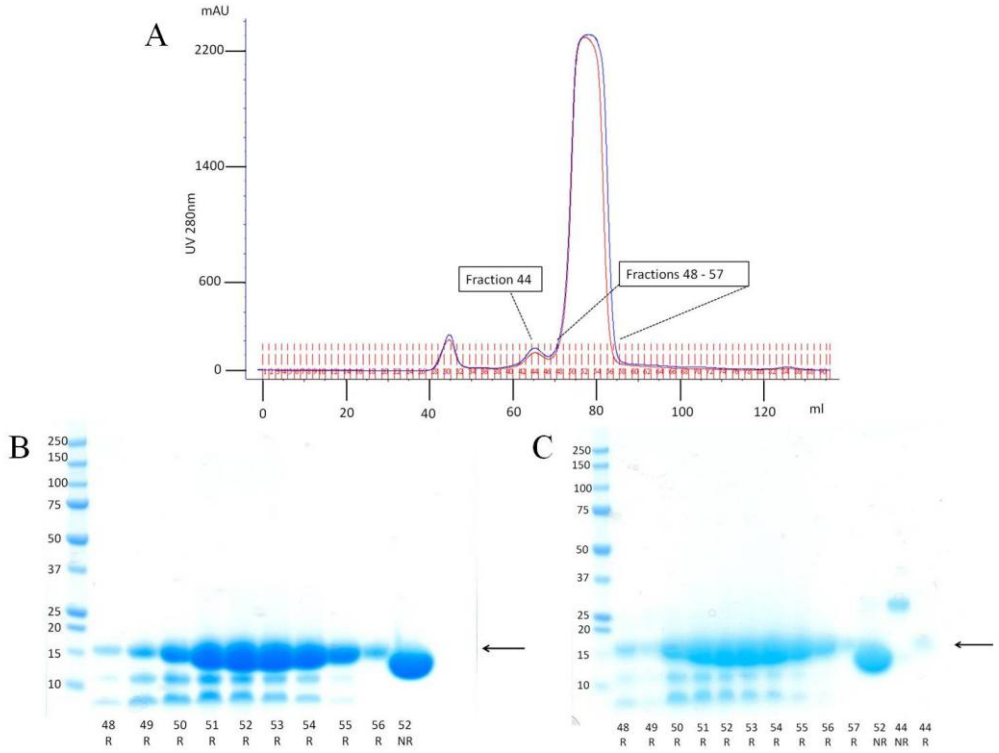


Figure 5.6 Size-exclusion chromatography of 12-13 wild type using HiLoad 16/60 Superdex 75. (A) Elution profiles (blue and red lines) overlaid. Gradient SDS-PAGE shows fractions from injection 1 (B) and injection 2 (C) in both reducing and non-reducing conditions. Bands of interest are indicated by arrows. The band visible at ~ 30 kDa in fraction 44 of (C) is probably dimeric FH 12-13.

Fractions 48-57 from both injections were combined and concentrated to 1.0 ml (500 µM). The sample was thought to be pure enough to record a ^{1}H, ^{15}N-HSQC so was dialyzed into 20 mM potassium phosphate, pH 6.6. The spectra (not shown) showed that the sample was not pure since it contained peaks that were consistent with degradation. In an attempt to improve its purity, the sample was subjected to a further size- exclusion chromatography step, using a Superdex 75 10/30 column.

An initial run on the Superdex 75 10/30 column contained 200 µl of FH 12-13 and 250 µM TCEP (included to try and remove some of the clipped protein from the sample; this concentration had been previously optimized by Dr Isabell Pechtl [162]). The elution profile (Figure 5.7, (A)) contained two peaks; a higher-molecular weight peak at around 67 kDa and a lower molecular weight peak at around 13.7 kDa (correct for FH 12-13) based on a calibration profile. Analysis of the two peaks by SDS-PAGE (Figure 5.7, (B)) showed that the material in the 67-kDa peak gave a smear in non-reducing conditions but under reducing conditions returned back to the expected size for monomeric FH 12-13. Clipped protein was still present in the 13.7-kDa peak. Fraction 9 was dialyzed back into 20 mM potassium phosphate. A sample at pH 6.6 and in 500 µl (at approx 200 µM protein, 2.8 mg/ml) was used to record the ^{1}H, ^{15}N-HSQC spectra for comparison against the original FH 12-13 wild type spectra[3] (see section 5.4). From a 1 L fermentation the approximate yield of ^{15}N labeled 12-13 WT is 3 mg/L.

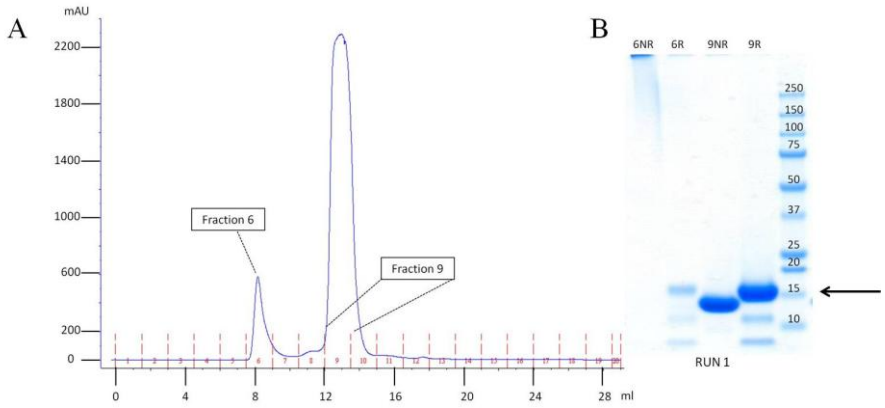


Figure 5.7 Purification of FH 12-13 wild type by chromatography on a Superdex 75 HR 10/30 size-exclusion column. The elution profile is shown in (A). In (B) is the gradient SDS-PAGE for the run, with the band suspected to correspond to FH 12-13 indicated by the arrow.

5.3.3 Purification of FH 12-13 3xGLY

The supernatant containing FH 12-13 3xGLY was diluted threefold and then adjusted to pH 5.0 prior to loading onto a 30 ml SP-Sepharose column. After loading the protein was eluted using a 1 M NaCl gradient. The elution profile (Figure 5.8) contained a single peak that was very small in comparison to FH 12-13 wild-type.

[3] Original spectra were recorded by Dr Christoph Schmidt as part of his PhD thesis.

This began to elute at approximately 300 mM NaCl. SDS-PAGE analysis of the peak in non-reducing conditions showed three weak bands: a high-molecular weight band of over 250 kDa, a band at approximately 25 kDa which might correspond to dimeric FH 12-13 3xGLY, and a band at 14 kDa which could be FH 12-13 3xGLY. In reducing condition the high-molecular weight band remains but the band thought to be a dimer is replaced by a band consistent with momomeric FH 12-13 3xGLY. As was seen in the expression of FH 12-13 wild type, the reduced sample for FH 12-13 3xGLY also contain lower bands likely corresponding to clipped versions of the protein.

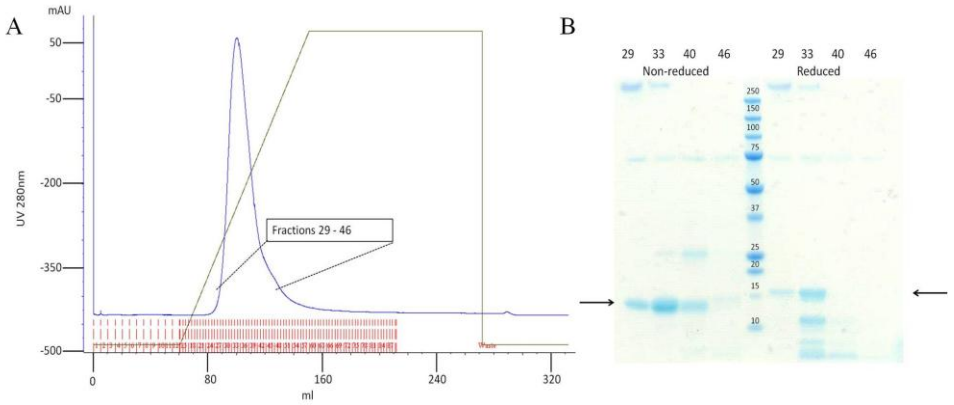


Figure 5.8 Purification of FH 12-13 3xGLY on SP-Sepharose. (A) The elution profile using a 1 M NaCl gradient. The blue line shows the UV absorbance at 280 nm and the green line shows the NaCl gradient (0-100%). (B) Gradient SDS-PAGE of fractions indicated on the chromatogram in both non-reducing and reducing conditions. The bands suspected to be FH 12-13 are indicated by arrows. The band visible at 75 kDa is Endo H_f.

As before it was important to conserve protein so protein-containing fractions (28-41) were combined to give a final volume of 30 ml. The fractions were treated with Endo H_f and then concentrated and buffer-exchanged into PBS to a final sample volume of 1.0 ml.

A single size-exclusion chromatography step was then performed (Figure 5.9). The chromatogram contains a main peak with two small peaks or shoulders present on either side. The peak/shoulder to the left of the main peak appears to be FH 12-13 3xGLY dimer as seen on the SDS-PAGE (fraction 44, non-reduced only) and the peak to the right was ignored. Analysis of the main peak by SDS-PAGE showed a single band in non-reducing conditions for fraction 52 but, under reducing

conditions, fractions 49-57 revealed the presence of a high proportion of clipped proteins in addition to some candidate FH 12-13 3xGLY.

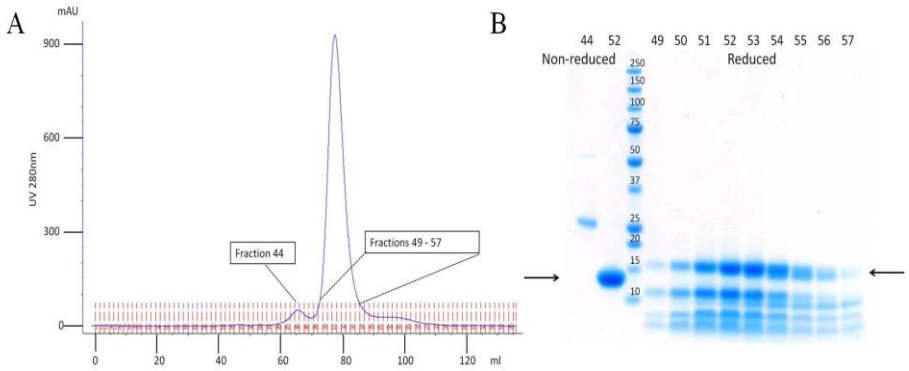


Figure 5.9 FH 12-13 3xGLY eluted from a HiLoad 16/60 Superdex 75 size exclusion chromatography column. (A) Elution profile. (B) The gradient SDS-PAGE gel shows analysis of the main protein-peak fractions.

Fractions 49-57 were then concentrated to 1.0 ml for the final purification step using the Superdex 75 10/30 size-exclusion column. The resulting elution profile (Figure 5.10 (A)) showed a single protein peak with a small leading-edge shoulder. Fractions 26-30 from the main peak were analyzed by SDS-PAGE (Figure 5.10 (B)) and under reducing condition contained the main protein band at 15 kDa and clipped proteins which appear to have decreased in amount compared to the previous purification step. The non-reduced sample prepared from fraction 28 contains the FH 12-13 3xGLY band at ~ 14 kDa but, in this heavily loaded sample, a slightly lower band is revealed that wasn't visible previously.

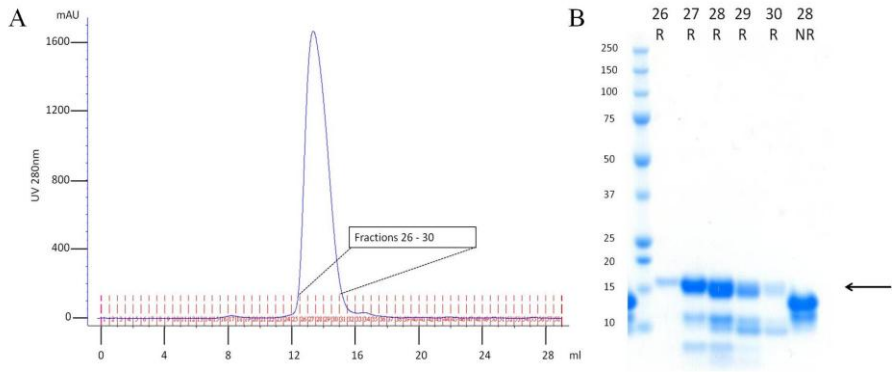


Figure 5.10 Purification of FH 12-13 3xGLY by Superdex 75 HR 10/30 size-exclusion chromatography. The chromatogram shown in (A) shows the main protein peak at 14 ml and the gradient SDS-PAGE for the fractions is shown in (B). The arrow indicates the protein bands.

The fractions were then combine and concentrated to give a final sample of 1.0 ml at approximately 100 µM, 1.4 mg/ml. An NMR sample was prepared by dialyzing 500 µl of the sample into 20 mM potassium phosphate, pH 6.6 to enable a ^{1}H, ^{15}N-HSQC spectra to be recorded for comparison with FH 12-13 wild type (see section 5.4).). From a 1 L fermentation the approximate yield of ^{15}N labeled 12-13 TT is1.4 mg/L.

5.3.4 Purification of FH 12-13 SL

After harvesting, the supernatant was treated and purified in the same way as the supernatant for FH 12-13 3xGLY. The elution profile showed one main peak (300 mM NaCl) with a leading-edge shoulder (Figure 5.11 (A)). SDS-PAGE analysis (Figure 5.11 (B)) of the fractions in both non-reducing and reducing conditions showed that the main peak contained the target protein, and the tail was most likely impurities. As was seen for both FH 12-13 3xGLY and FH 12-13 wild type, the reduced samples for FH 12-13 SL also contain clipped versions of the protein.

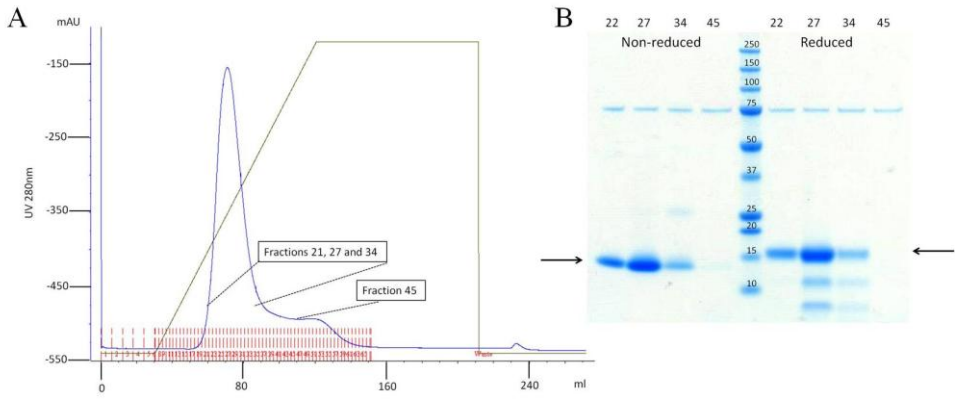


Figure 5.11 Purification of FH 12-13 SL on SP-Sepharose. (A) Elution profile (blue line). The green line is the NaCl gradient (0-100%). (B) The fractions indicated on the profile were analyzed by SDS-PAGE, which showed that protein of the expected size (arrow) was found only in the main peak.

Fractions 22-27 were combined and then the sample was treated with Endo H_f prior to concentration and buffer-exchange into PBS. A final volume of 1.0 ml was achieved, which was then split into two 0.5 ml samples for size-exclusion chromatography (Figure 5.12). The separate injections both gave the same result of a single peak eluting at 80 ml. Analysis of the fractions from the peak by SDS-PAGE

showed that under non-reducing conditions there is a single band at the expected size, with the lower-molecular weight clipped contaminants still present. Fractions 52-56 from both injections were combined and concentrated to a final volume of 1.0 ml for a final purification step using the Superdex 75 10/30 size-exclusion chromatography column.

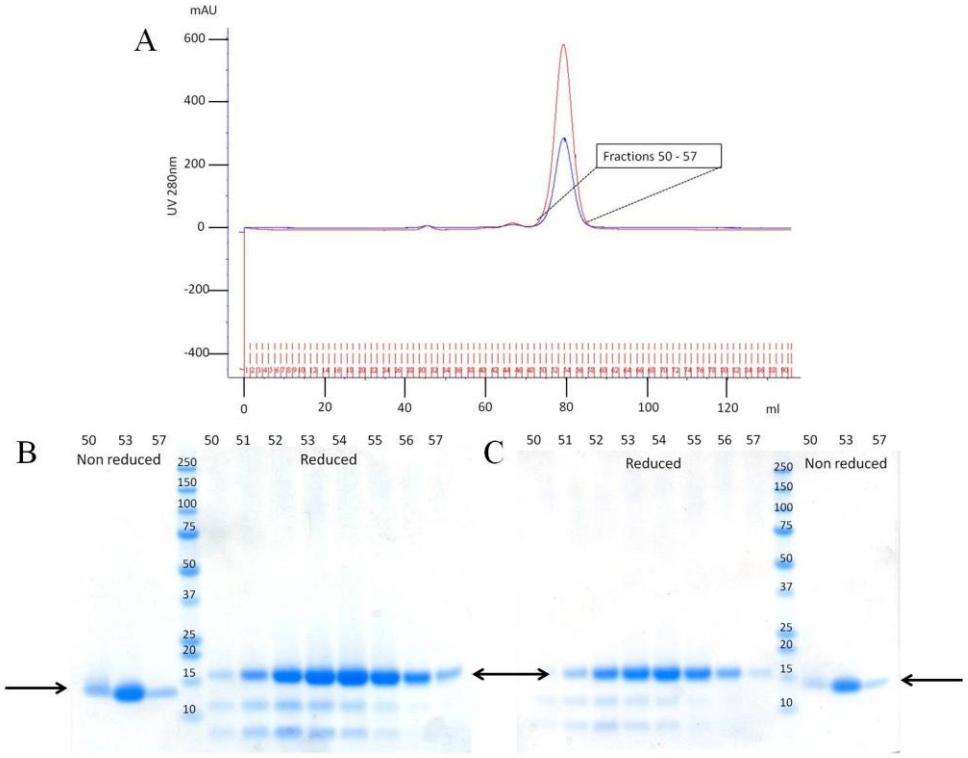


Figure 5.12 Size-exclusion chromatography of 12-13 SL using HiLoad 16/60 Superdex 75. (A) Elution profiles (blue and red lines) containing the main protein peak for both injections overlaid. Gradient SDS-PAGE was performed for fractions from injection 1 (B) and injection 2 (C) in both reducing and non-reducing conditions. The bands of interest are indicated by arrows.

Following size-exclusion chromatography on Superdex 75 10/30, the elution profile (Figure 5.13 (A)) contained a peak at 14 ml, corresponding to approximately 14 kDa (based on a previous calibration of the column). Analysis by SDS-PAGE (Figure 5.13, (B)) (in reducing conditions only) showed only a trace of clipped proteins although the gel was not heavily loaded.

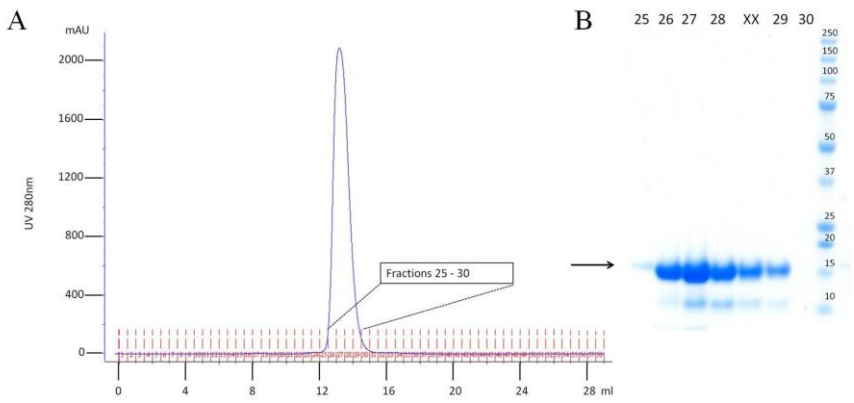


Figure 5.13 Purification of FH 12-13 SL by size-exclusion chromatography on Superdex 75 HR 10/30. (A) Elution profile. (B) The gradient SDS-PAGE for the fractions, in reducing condition only. The lane labeled XX is FH 12-13 3xGLY, which was loaded in error.

The fractions (26-39) were combined and concentrated to a final volume of 1.0 ml and concentration of ~200 µM, 2.8 mg/ml which was then dialyzed into 20 mM potassium phosphate pH 6.6. A ^{1}H, ^{15}N-HSQC was recorded with 500 µl of the sample for comparison with FH 12-13 wild type (see section 5.4).). From a 1 L fermentation the approximate yield of ^{15}N labeled 12-13 SL is 3 mg/L.

5.3.5 Purification of FH 8-15 wild type

Production of FH 8-15 WT was carried out in shaker flasks (see Methods, section 2.2.5). After removal of cells the resulting supernatant was diluted six-fold and adjusted to pH 8.5 for application to a 50 ml Q-Sepharose column. The protein was eluted using a 1 M NaCl gradient with the elution profile (figure 5.14 (A)) containing multiple overlapping peaks. Analysis by SDS-PAGE (figure 5.14 (B)) confirmed that a protein that had eluted between 100 and 200 mM NaCl ran at the expected size of ~50 kDa. The contaminants were thought to be proteins from *P. pastoris* and the yield of target protein was very low indeed compared to FH 12-13.

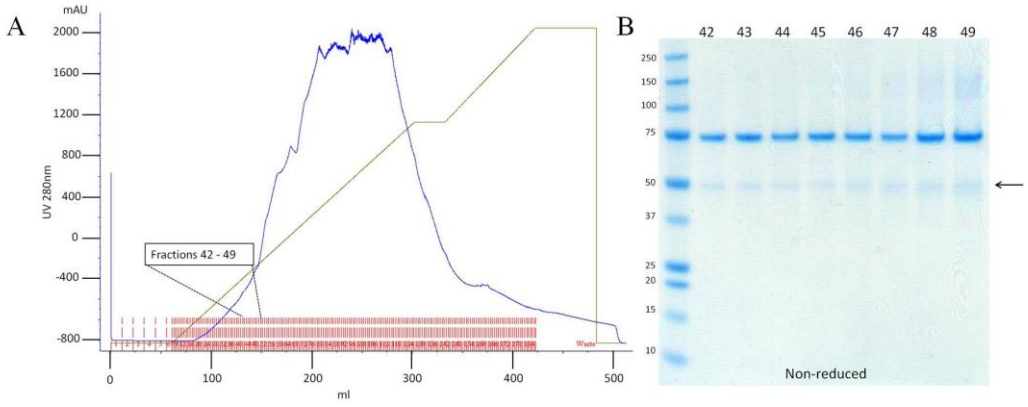


Figure 5.14 Elution of FH 8-15 wild type from Q-Sepharose column. (A) Elution profile. The blue line is the UV absorbance at 280 nm and the green line indicates the NaCl gradient (0-50% over 6 CV, 50% for 1 CV and 50-100% over 2 CV). Fractions indicated on (A) were then analyzed by SDS-PAGE in non-reducing conditions (B) with the putative band of interest (judging by its apparent MWt) indicated by the arrow. The band visible at 75 kDa is Endo H_F.

Despite the very low yield obtained, the protein-containing fractions 42-49 were then pooled (approx 40 ml) and the pH was adjusted to 6.0 for de-glycosylation with Endo H_f. After treatment, the protein was concentrated to a final volume of 1.5 ml and buffer exchanged into 20 mM potassium phosphate with 500 mM NaCl, pH 7.4 for application to a Superdex 200 10/300 GL size-exclusion chromatography column. Three separate injections of ~ 500 µl were performed and fractions from runs 1 and 2 were then analyzed by SDS-PAGE (Figure 5.15).

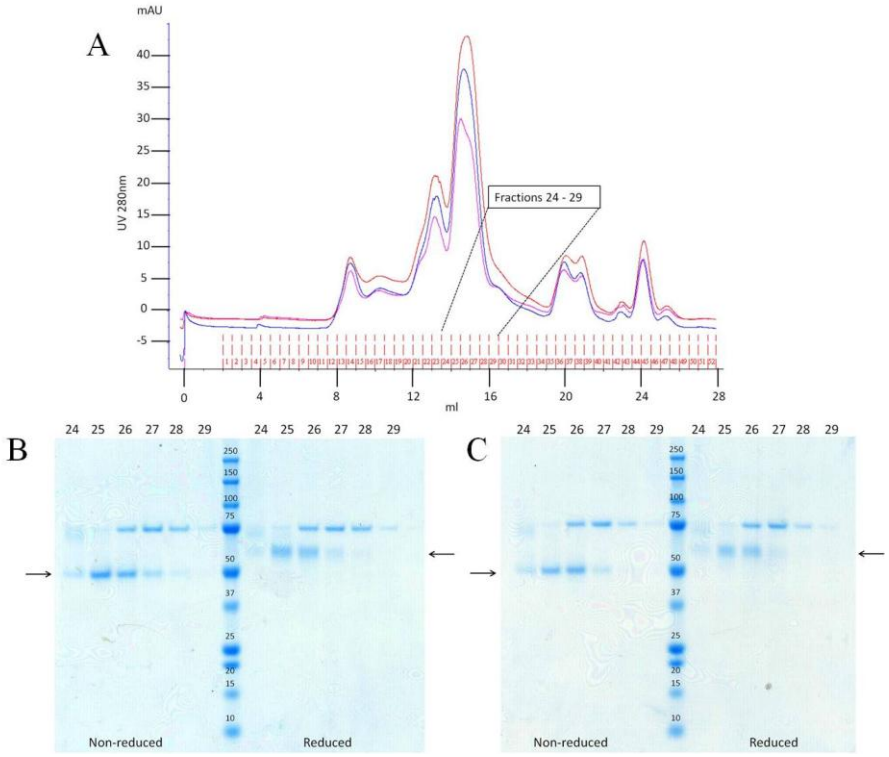


Figure 5.15 Size-exclusion chromatography of FH 8-15 WT, using a Superdex 200 HR 10/30 column. (A) The three elution profiles are shown overlaid while the analysis by gradient SDS-PAGE of the fractions indicated for run 1 is shown in (B) and for run 2 is shown in (C).

Analysis by SDS-PAGE showed that a protein of the expected size eluted between 14 ml and 16 ml (fractions 24-29). In all fractions a band was visible at 75 kDa. This cannot be excess Endo H_f since after treating fraction 26 with Amylose resin the band was still present (not shown). It could be the AOX1 protein which is expressed alongside the protein of interest and when in monomer form has a molecular weight of 73.9 kDa [176]. In order to try to remove this larger protein from the FH 8-15 WT preparation, fractions 24-26 from all runs were concentrated and buffer-exchanged into 20 mM sodium carbonate buffer, pH 9.0, for application to a Tricorn Mono Q column. The bound proteins were then eluted over 20 column volumes to a final concentration of 500 mM NaCl followed by three column volumes to reach 1 M NaCl (100%). Finally, SDS-PAGE was used to analyze the fractions (Figure 5.16).

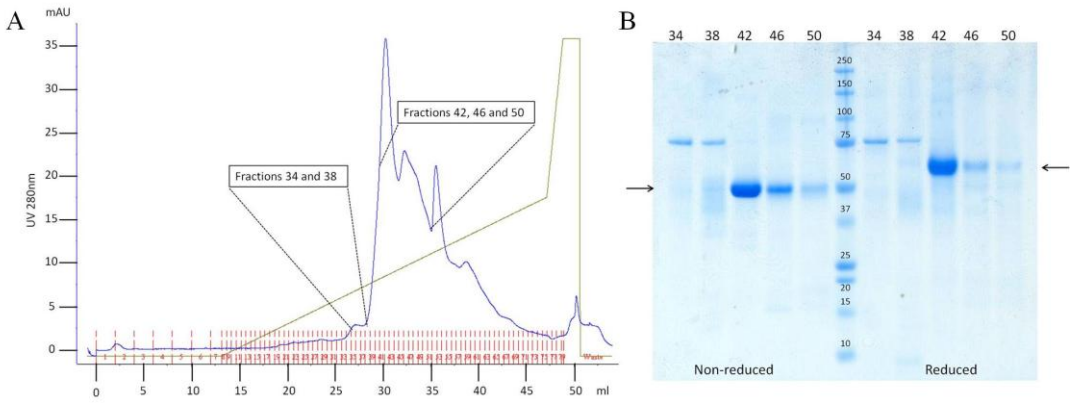


Figure 5.16 FH 8-15 WT Mono Q. (A) Elution profile and (B) gradient SDS-PAGE for the fractions highlighted on (A). In (A) the blue line is the UV absorbance at 280 nM and the green is the concentration of NaCl (0-50% over 20 CV and 50-100% over 3 CV).

The elution profile showed five peaks eluting between 200 and 400 mM NaCl and analysis of the peaks showed that the 75-kDa protein was eluted first at 27 ml and 8-15 WT eluted subsequently, between 30 and 35 ml. Fractions 42-47 were pooled and stored at -80 °C. As the concentration of protein was too low (2.5 µM, 0.01 mg/ml in 1 ml) for SAXS experiments, another sample was prepared (14.2 µM, 0.77 mg/ml in 1 ml (Figure 5.17). For SPR studies (see chapter 6) the final concentration was 6.4 µM (0.34 mg/ml) in 1 ml.

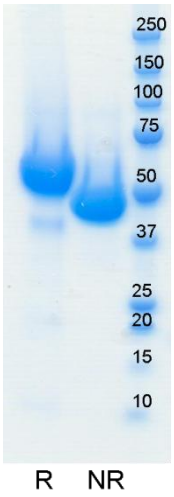


Figure 5.17 Final FH 8-15 WT sample for SAXS. Bands appear at the expected molecular weight though under reducing conditions a small amount of degradation is visible.

5.3.6 Purification of FH 8-15 3xGLY

A 10-L fermentation was preformed to produce FH 8-15 3xGLY. A final volume of 7.7 L of supernatant (after removal of cells) was adjusted to pH 4.0 and diluted fivefold for application to a 50 ml SP Sepharose cation-exchange column. The bound protein was eluted using a 1 M NaCl gradient and the elution profile (Figure 5.18 (A)) showed one very large broad peak eluted between 400 and 900 mM NaCl.

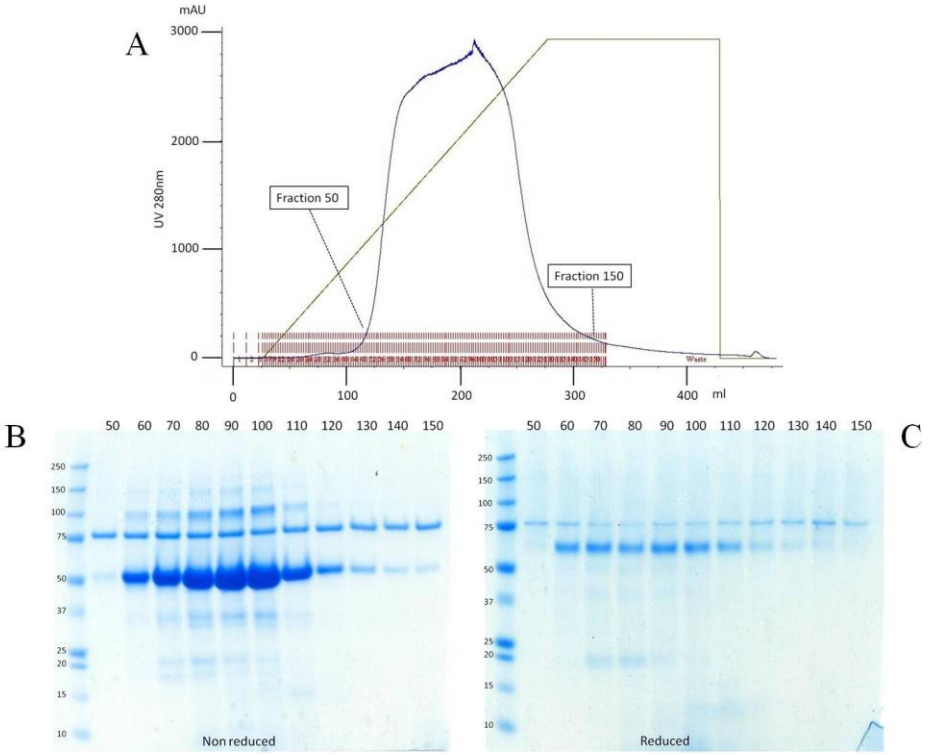


Figure 5.18 FH 8-15 3xGLY elution from the SP Sepharose column. (A) Elution profile (blue line). The green line indicates the NaCl gradient (0-100%). (B) The gradient SDS-PAGE shows fractions 50-150 in non-reducing conditions; (C) shows the fractions in reducing conditions. The band visible at 75 kDa is Endo H_f.

Analysis of the fractions by SDS-PAGE showed that protein of the expected sizes was present in fractions 60-140 but these contained several impurities including possible dimer (100-kDa band in non-reducing conditions). Fractions 50-120 (approx 200 ml) were pooled and after adjusting pH to 5.6 underwent Endo H_f treatment. The next step was to concentrate 50 ml down to ~ 3 ml and buffer exchange into 20 mM sodium carbonate, pH 9, for anion-exchange chromatography (the remaining protein was frozen). Three separate injections (1.0 m, 1.0 ml and 0.5

ml injections) were performed and the column was eluted using the same program as 8-15 WT.

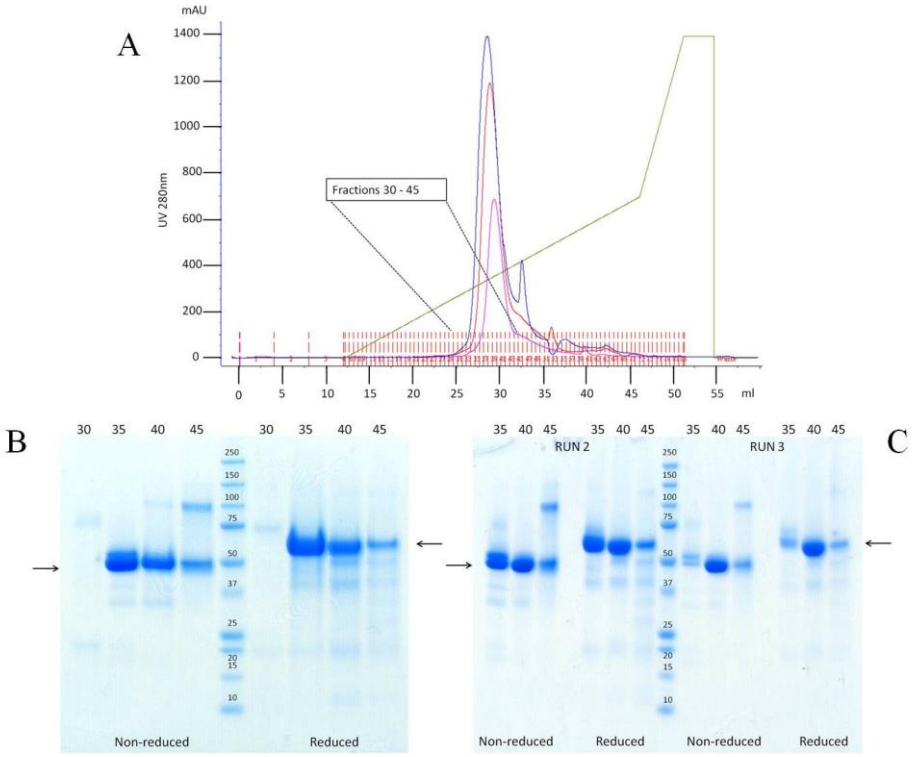


Figure 5.19 The 8-15 3xGLY construct eluted from a Mono Q column. (A) Elution profiles for all three injections overlaid. The green line is the concentration of NaCl (0-50% over 20 CV and 50-100% over 3 CV). The gradient SDS-PAGE in (B) shows the fraction indicated on (A) for injection 1 whilst (C) shows the fractions from injections 2 and 3. Both gels were run in non-reducing and reducing conditions.

The elution profiles (Figure 5.19 (A)) each showed the target protein peak eluting between 25 and 33 ml (200 and 300 mM NaCl) followed by smaller peaks that contained impurities (gel not shown). Analysis of injection 1 fractions by SDS-PAGE (Figure 5.19 (B)) showed that fraction 30 contained the 75-kDa and 20-kDa impurities so were discarded. Fractions 35 and 40 were mainly 8-15 3xGLY but still contained a small amount of impurities (~ 37 kDa) and fraction 40 still contained some dimer protein (100 kDa). Injections 2 and 3 (Figure 5.19 (C)) showed identical results to injection 1. Fractions 33-40 were pooled from all injections and buffer exchanged into 20 mM potassium phosphate buffer, pH 7.4, with 500 mM NaCl prior to being concentrated to a final volume of 1 ml for size-exclusion chromatography.

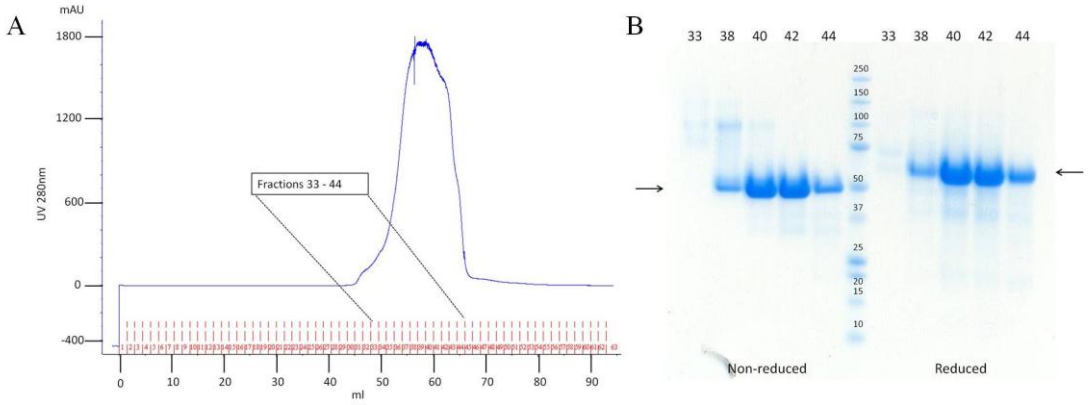


Figure 5.20 Size-exclusion chromatography of 8-15 3xGLY using HiLoad 16/60 Superdex 75. (A) Elution profile. (B) The gradient SDS-PAGE of the fractions indicated on (A). (elution profile was incomplete due to a power cut).

A broad asymmetrical peak was seen in the elution profile (Figure 5.20 (A)) eluting between 45 and 70 ml. Analysis of the fractions by SDS-PAGE indicated that fraction 33 was mostly dimer (100 kDa) which was also present (~ 5-10%) in fractions 38 and 40. Fractions 42 and 44 contained pure 8-15 3xGLY so along with Fractions 41 and 43 were pooled. The final concentration of the sample was 21.6 µM in 1 ml and this sample was used for SAXS analysis

The purification process also generated a sample for SPR binding studies (see chapter 6) at 11 µM (0.5 mg/ml) in 1.5 ml.

5.3.7 Purification of FH 8-15 SL

As for 8-15 3xGLY, a 10-liter fermentation was carried out to produce 8-15 SL. The resulting supernatant (7 L) was adjusted to pH 4, diluted six-fold and then the purification protocol used for 8-15 3xGLY was followed.

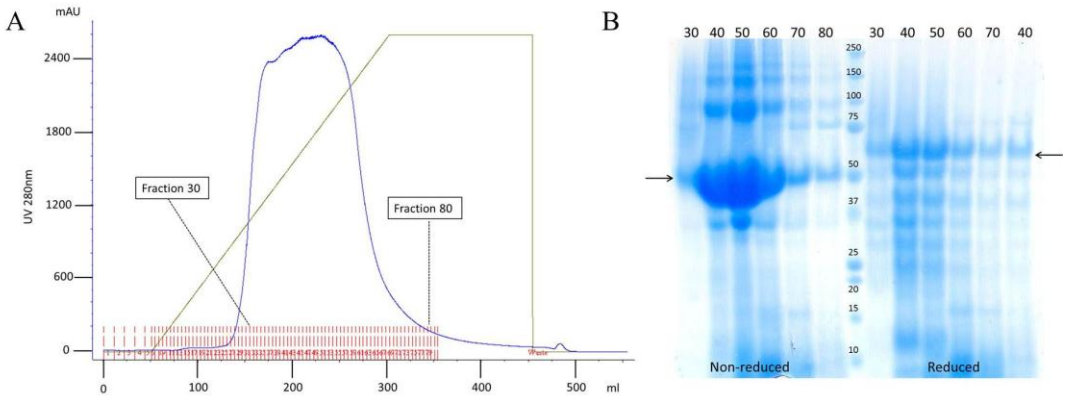


Figure 5.21 Elution of 8-15 SL from SP Sepharose. (A) Elution profile. The blue line is the UV absorbance at 280 nm and the green line indicates the NaCl gradient (0-100% over 4 CV). (B) The gradient SDS-PAGE for the fractions from the elution. Lane number corresponds to the fraction number. Target protein bands are indicated by arrows.

The protein that had bound to the column eluted between 400 and 800 mM NaCl (150-300 ml). Analysis of the fractions by SDS-PAGE revealed that most of the target protein had eluted in fractions 40-60. These fractions were pooled, deglycosylated and 30 ml (out of 120 ml) was buffer exchanged for application to a Mono Q anion-exchange chromatography column. Figure 5.22 shows the elution profile and SDS-PAGE for one of the injections (3 ml loaded out of 30 used, each injection ~ 500 µl).

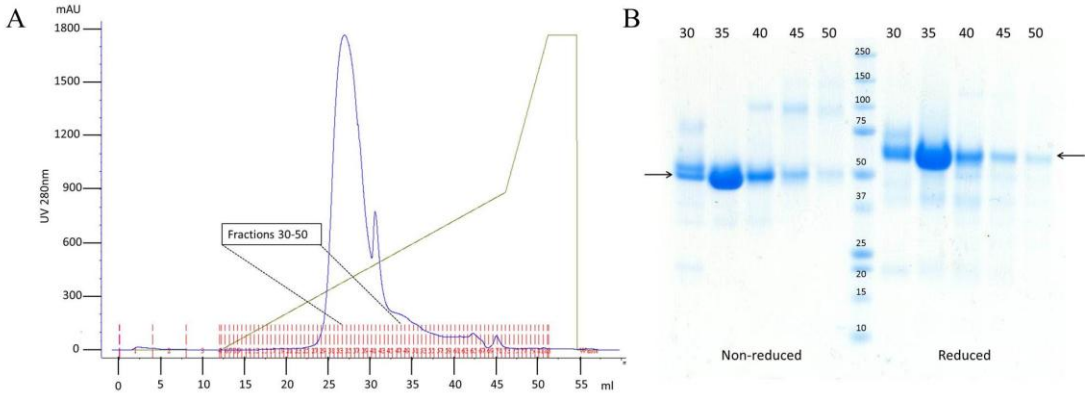


Figure 5.22 Elution of FH 8-15 SL from the Mono Q column. (A) The elution profile (blue line); the green line is the concentration of NaCl (0-50% over 20 CV and 50-100% over 3 CV). (B) The gradient SDS-PAGE for the fractions indicated on (A).

Two overlapping peaks were visible on the elution profile (Figure 5.22 (A)) and SDS-PAGE analysis (Figure 5.22 (B)) revealed that fraction 30 gave a double band (the expected one at ~50 kDa, and one running slightly higher) under non-reducing conditions, which under reducing conditions appears as a single band suggesting that the higher band is a modified version of 8-15 SL. Fractions 40-50 contained dimer. Fractions 33-37 were buffer exchanged and concentrated to 1.0 ml to allow two injections of 0.5 ml onto a size-exclusion chromatography (Figure 5.23).

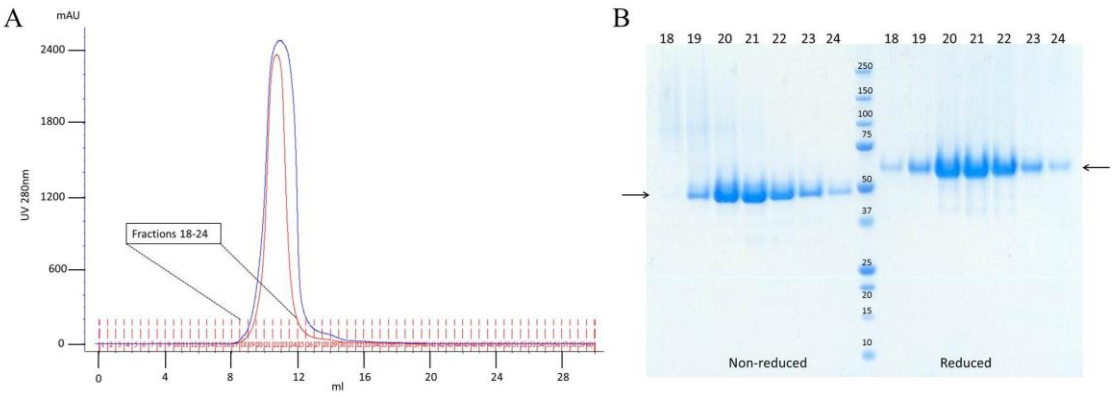


Figure 5.23 Size exclusion chromatography of FH 8-15 SL using a Superdex 200 HR 10/30 column. (A) Two profiles overlaid. (B) The gradient SDS-PAGE of the fractions indicated on (A) for injection one.

Both injections showed a single peak eluting between 8 and 13 ml. The fractions form the first injection showed that 18-19 still contained a very small amount of dimer and fractions 20-24 contained the cleanest protein. Fractions 18-19 from both injections were pooled as were fractions 20-24. The latter fractions (20-24) were concentrated (to 30 µM, 1.5 mg/ml in 1.0 ml) to provide a sample for SAXS analysis

The purification process was repeated using the remaining 27 ml (remaining from the initial Mono Q chromatography step) to generate a sample for SPR binding studies at a final concentration of 4.5 µM (0.2 mg/ml) in 9 ml.

5.3.8 Purification of FH 10-15 WT

FH 10-15 WT production was carried out in shaker flasks (see Methods, section 2.2.5). The supernatant was adjusted to pH 6 and diluted fivefold for application to a

30 ml SP Sepharose column (GE Healthcare). Bound protein was then eluted using a 1 M NaCl gradient.

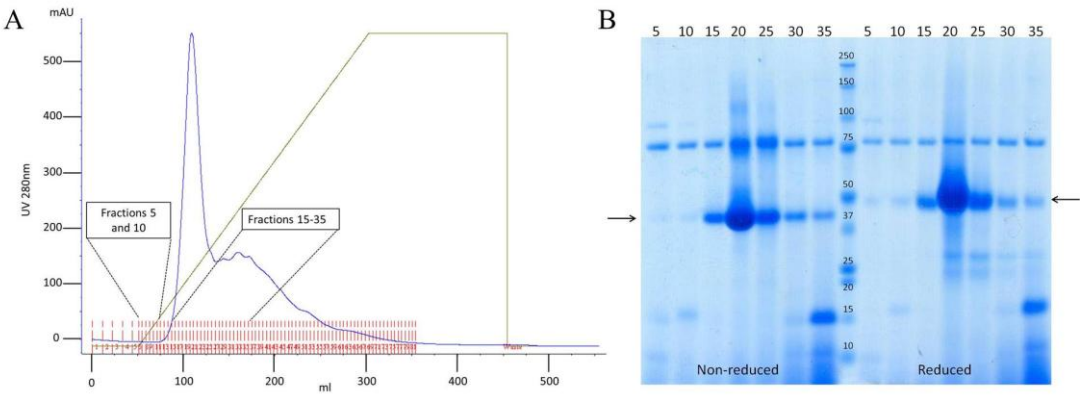


Figure 5.24 10-15 WT SP Sepharose elution. (A) Elution profile. The blue line is the UV absorbance at 280 nm and the green line indicates the NaCl gradient (0-100% over 4 CV). (B) Gradient SDS-PAGE gel of the fractions indicated on (A). Target protein bands are indicated by arrows.

The elution profile (figure 5.24 (A)) showed a small sharp peak eluting between 100 200 NaCl followed by many overlapping peaks. The SDS-PAGE analysis of the fractions (Figure 5.24 (B)) showed that the protein was present mainly in fraction 20 that corresponds to the top of the main peak. Fractions 15-25 were pooled, deglycosylated and buffer exchanged prior to concentration to 1.5 ml for anion-exchange chromatography. Three injections of 0.5 ml were performed.

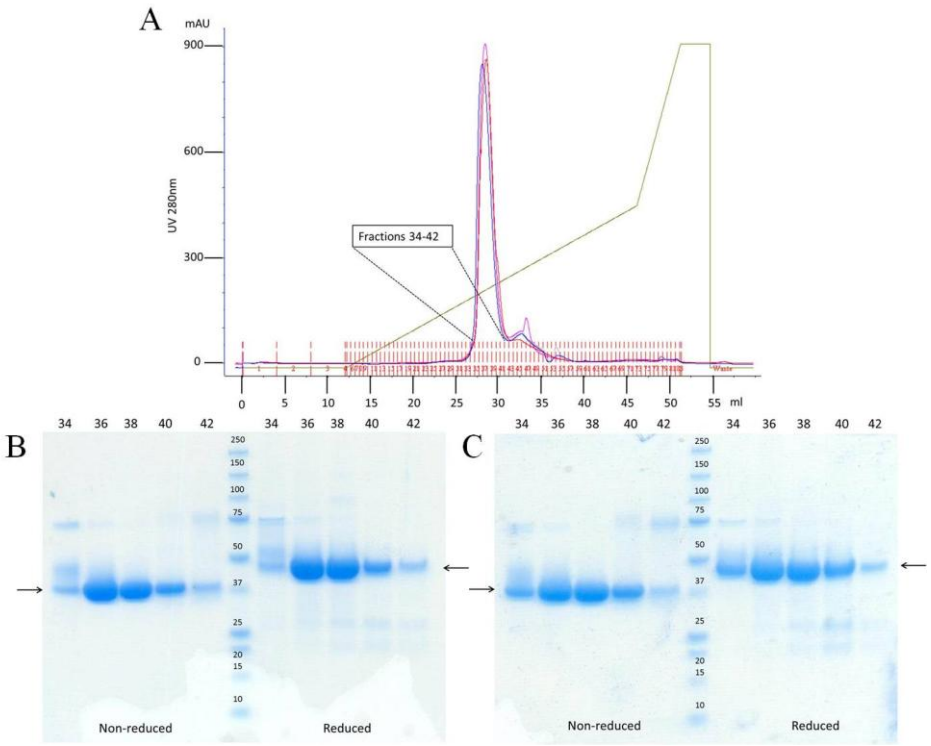


Figure 5.25 FH 10-15 WT eluted from a Mono Q column. The injection profiles (blue and red) are shown overlaid in (A) whilst the gradient SDS-PAGE gels show the fractions indicated with (B) corresponding to the first injection and (C) to the second.

Both injections showed a sharp peak eluting between 200 and 250 mM NaCl (Figure 5.25 (A)) and analysis of the fractions by SDS-PAGE for the first and second injections showed that fractions 36-40 contained nearly all the protein (Figure 5.25 (B)). Fraction 34 was discarded from each run due to the diffused band seen above the protein band whilst fractions 36-40 were pooled, buffer exchanged and concentrated to 1.0 ml for size-exclusion chromatography.

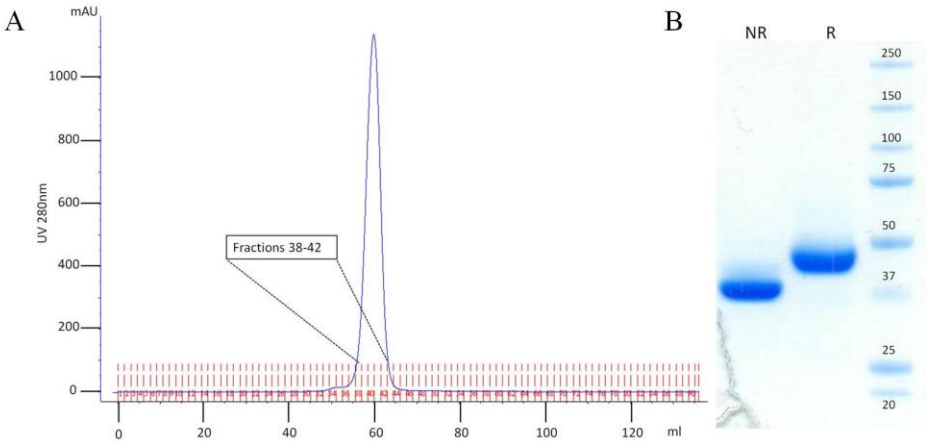


Figure 5.26 Size-exclusion chromatography of FH 10-15 WT using HiLoad 16/60 Superdex 75. (A) shows the elution profile and (B) shows the final purified sample.

A single peak was eluted between 55 and 65 ml (Figure 5.26 (A)) and when fractions 38-42 were analyzed by SDS-PAGE all contained clean protein (data not shown). The fractions were then pooled and concentrated, and as a final purity check, 5 µl was loaded onto SDS-PAGE (Figure 5.26 (B)).

The sample used for SAXS analysis was 43 µM (1.8 mg/ml) in a final volume of 1.0 ml). From 1 L of BMMY media the approximate yield is is 2 mg/L.

5.3.9 Purification of FH 10-15 3xGLY

A 10-litre fermentor was used to express 10-15 3xGLY. The final volume of supernatant (after cell removal) was 6.3 L and purification was carried out as described for 8-15 3xGLY (see section 5.36 above). Elution of the SP Sepharose column showed a broad peak eluting between 350 mM and 1 M NaCl (Figure 5.27 (A)). Analysis of the fractions by SDS-PAGE (Figure 5.27 (B) and (C)) showed that fractions 60-150 contained the protein at the expected sizes. Fractions 70-150 were combined (volume ~200 ml) for Endo H_f digestion.

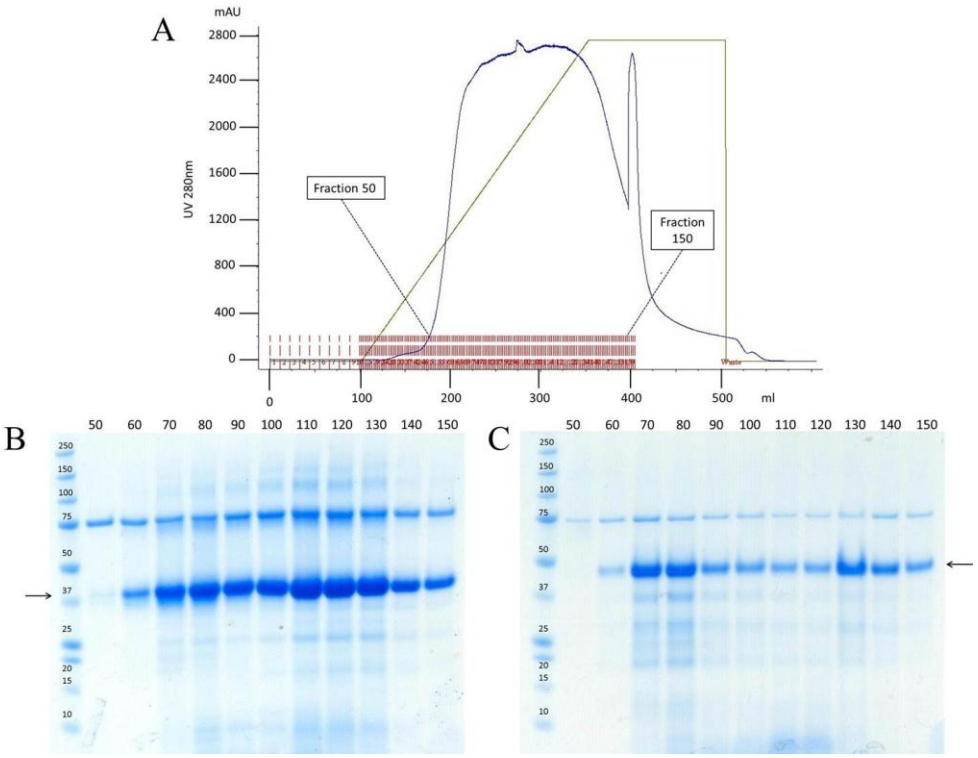


Figure 5.27 Elution of FH 10-15 3xGLY from SP Sepharose. (A) The elution profile (blue line). The green line is the NaCl gradient. (B) and (C) Gradient SDS-PAGE gels show the fractions indicated on (A) in non-reducing conditions (B), and reducing conditions (C).

After digestion, 50 ml was concentrated and buffer exchanged for anion-exchange chromatography (Figure 5.28) whilst the remaining ~150 ml was frozen and stored for future work. Three separate injections of between 0.5 and 1.0 ml were performed.

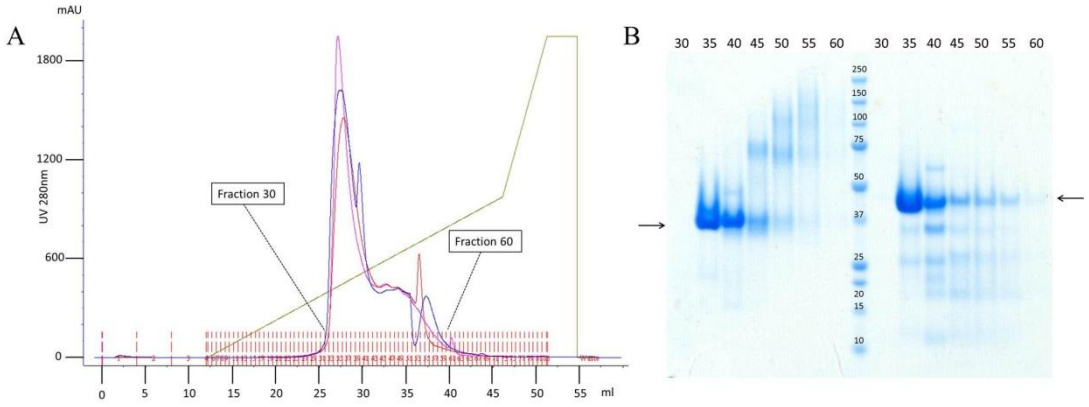


Figure 5.28 Elution of 10-15 3xGLY from Mono Q column. (A) The three elution profiles overlaid. (B) Gradient SDS-PAGE of injection one for the fractions indicated on (A).

The three injections each showed a peak eluting between 200 and 300 mM NaCl (25-30 ml) followed by a shoulder (30-40 ml). The SDS-PAGE showed that fractions 35-40 contained the protein and fractions 45-55 contained mainly dimer or trimer. Fractions 33-39 from each injection were pooled and prepared for size-exclusion chromatography (Figure 5.29). A final volume of 1.5 ml was injected onto the column.

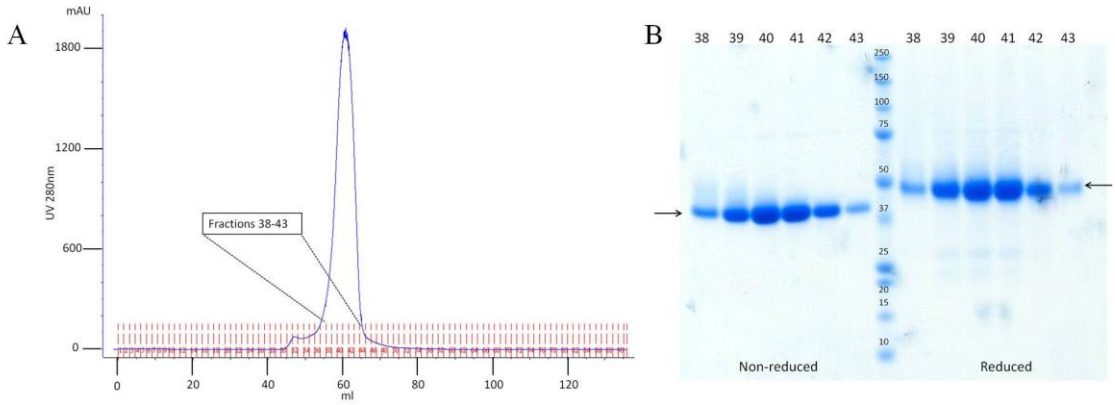


Figure 5.29 Size-exclusion chromatography of FH 10-15 3xGLY using HiLoad 16/60 Superdex 75. (A) Elution profile. (B) Gradient SDS-PAGE for the fractions indicated on (A).

A single main peak was seen between 55-65 ml and analysis of the fractions showed a single band under non-reducing conditions, but under reducing conditions a trace amount of clipped protein was detected.

For SAXS analysis the sample was 39.8 µM (1.62 mg/ml) in a final volume of 1.0 ml.

5.3.10 Purification of FH 10-15 SL

Expression and purification of 10-15 SL was carried out in an identical manner to that of 10-15 3xGLY. The elution profile from the SP Sepharose column (Figure 5.30 (A)) again showed a broad peak eluting between 400 mM and 1 M NaCl. Analysis of the fractions by SDS-PAGE (Figure 5.30 (B)) showed a protein band at the expected position and in reducing conditions revealed a high proportion of lower-molecular weight proteins. Fractions 20-44 were pooled (~200 ml) and treated with Endo H_f. As before, a 50 ml portion of this was concentrated and buffer exchanged

for anion-exchange chromatography while the remaining protein was frozen for future use.

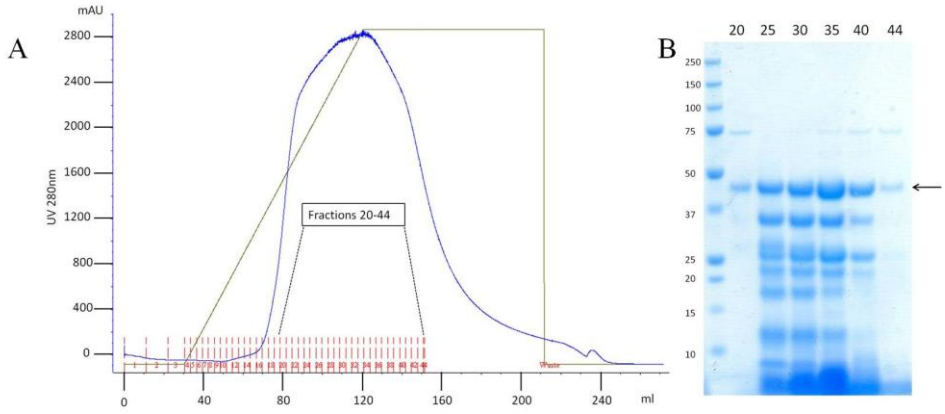


Figure 5.30 Elution of FH 10-15 SL from SP Sepharose. (A) Elution profile (blue line) is shown with the NaCl gradient indicated by the green line. (B) Gradient SDS-PAGE in reducing conditions for the fractions indicated on (A).

Three 1 ml injections were performed (only the data for the first injection is shown). The elution profile showed the main protein peak, with a shoulder, eluting between 200 and 300 mM NaCl and a smaller peak eluting later (400 mM NaCl). Analysis by SDS-PAGE confirmed that the fractions from the main peak contained the target protein whilst the smaller peak contained mainly dimer (75 kDa). Fractions 30-45 were pooled from each injection, buffer exchanged and concentrated to 1.5 ml for size-exclusion chromatography.

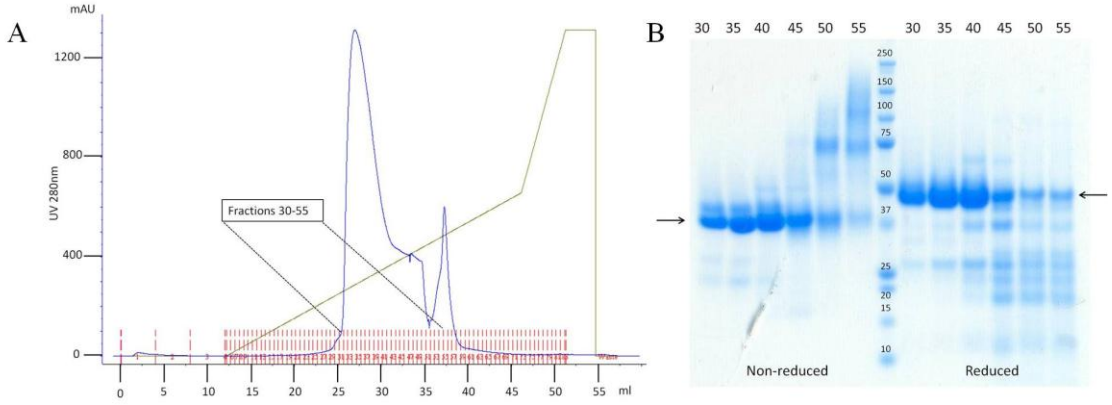


Figure 5.31 FH 10-15 SL eluted from the Mono Q column. (A) The elution profile (blue line) and NaCl gradient (green line). (B) The gradient SDS-PAGE from injection one for the fractions indicated on (A).

For size-exclusion chromatography, two injections of ~0.75 ml were performed. Each injection gave a single peak eluting between 9 and 13 ml (Figure 5.32 (A)). The SDS-PAGE for the fractions indicated showed that the protein was clean (gel not shown) and after being pooled and concentrated, 5 µl was run under non-reducing and reducing conditions as a final purity check (Figure 5.32 (B)).

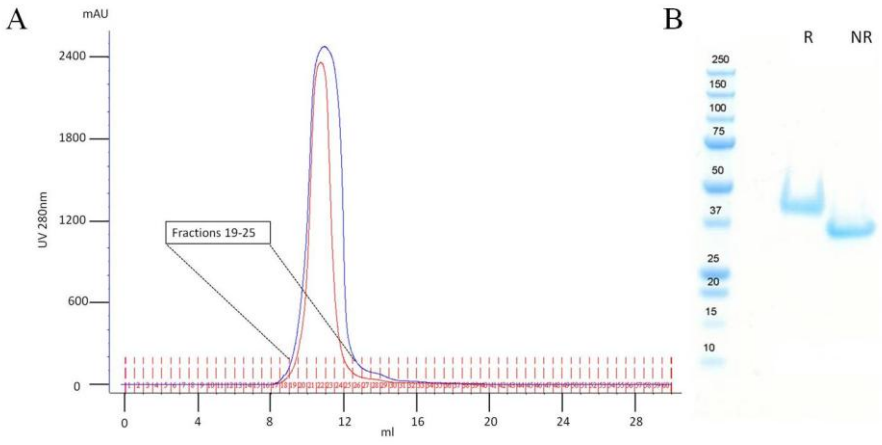


Figure 5.32 Size-exclusion chromatography of FH 10-15 SL on Superdex 75 HR 10/30. (A) Two elution profiles (red and blue) overlaid; (B) SDS-PAGE for the final check of purity.

The final concentration of the sample for SAXS analysis was 44 µM (1.78 mg/ml) in a final volume of 1.0 ml.

5.3.11 Purification of FH SL

For production of FH SL, a ten-litre fermentation was performed resulting in 7.4 L of supernatant after removal of cells. The supernatant was then diluted fivefold and applied to a 30 ml SP Sepharose equilibrated with 20 mM potassium phosphate plus 50 mM NaCl, pH 6.0. Bound protein was eluted using a salt gradient with the elution buffer being 20 mM potassium phosphate plus 1 M NaCl, pH 6.0. The elution profile (Figure 5.33 (A)) showed two overlapping peaks with a trailing-edge shoulder eluting between 100 and 900 mM NaCl. Analysis of the fractions by SDS-PAGE indicated that a protein of the expected size had eluted between fractions 17 and 21, although these fractions also contained impurities; fractions 17-20 were pooled (final volume 36 ml) and treated with Endo H_f.

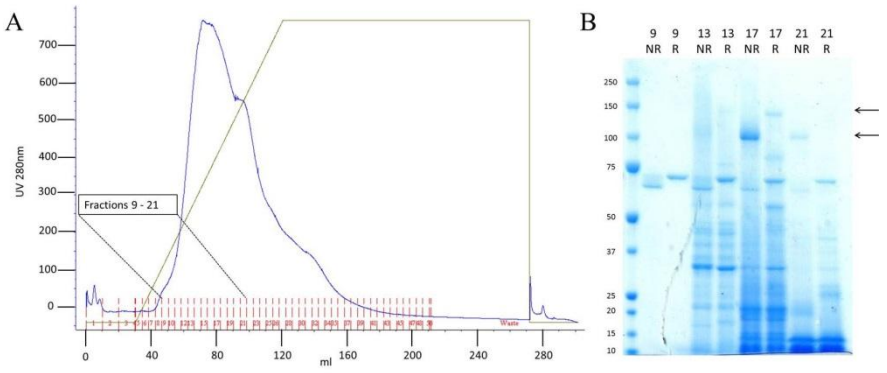


Figure 5.33 Elution of FH SL from SP Sepharose. (A) Elution profile using a 1 M NaCl gradient (0-100% over 3 CV, green line). (B) The gradient SDS-PAGE for the fraction indicated on (A), the upper arrow marks the expected position for reduced protein (R) and the lower for non-reduced (NR).

After Endo H_f digestion, the pooled fractions were passed over a 1.0 ml Concanavalin A (Con A, Ge Healthcare) bench-top column to remove the digested sugars (and any protein that was not deglycosylated). The resulting flow-through was then concentrated to 5.0 ml for application to a size-exclusion chromatography column. Multiple injections of between 0.5 and 0.7 ml were performed and the elution profiles are shown overlaid in Figure 5.34 (A).

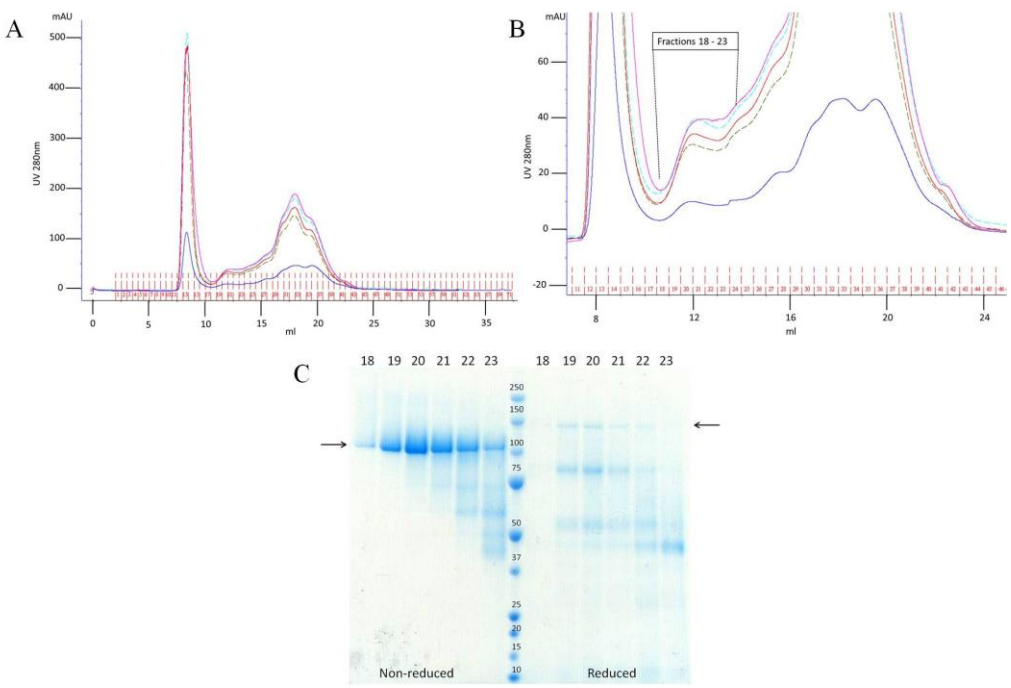


Figure 5.34 Sample of FH SL purified by size-exclusion chromatography using a Superdex 200 HR 10/30 column. (A) The overlaid profiles. (B) A zoomed-in view of (A) highlighting where FH SL eluted. (C) The gradient SDS-PAGE-based analysis of the fractions indicated on (B).

Multiple peaks were eluted. The peak eluting at 8-10 ml (fractions 12-16) contained no protein whilst the overlapping peaks between 15 and 20 ml contained the lower-molecular weight proteins seen on the SP elution gel when analyzed by SDS-PAGE. The SDS-PAGE suggested that FH SL eluted between 11 and 13 ml (fractions 18-23). Though the non-reduced samples gave only a very promising single band with a small amount of contaminant in fraction 22 and 23, the reduced samples showed that while FH SL was present there was, unfortunately, a large percentage of clipped protein.

Fractions 18-20 contained the least clipped protein, so these were pooled and concentrated to a final volume of 1.5 ml with a concentration of 0.87 µM. Fractions 21-23 were frozen and stored separately.

5.3.12 Purification of FH 3xGLY

Production and purification of FH 3xGLY was carried out via an identical procedure to that used for FH SL. After cell removal the final volume of supernatant was 7.8 L. The profile for the SP Sepharose elution gave a single peak eluting between 300 and 600 mM NaCl. Analysis of the fractions by SDS-PAGE strongly suggested that the target protein had eluted in fractions 15-21 but contained many lower-molecular weight impurities.

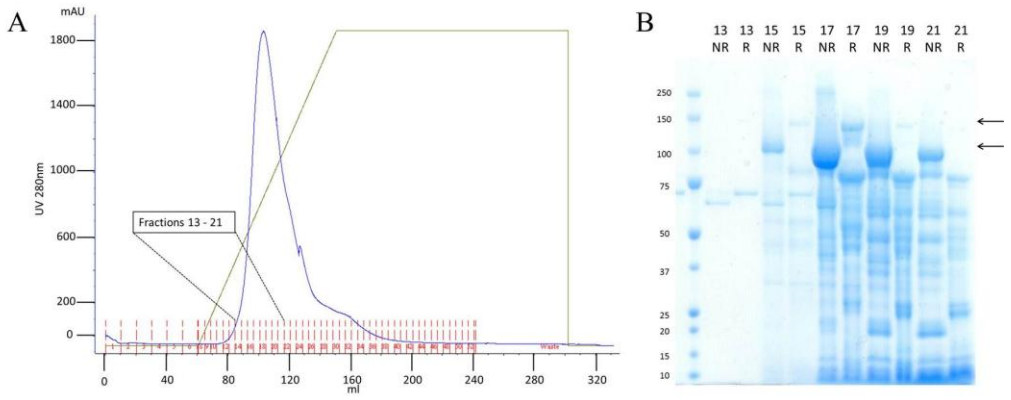


Figure 5.35 Elution of FH 3xGLY from SP Sepharose. (A) The elution profile (blue line) using a 1 M NaCl gradient (0-100% over 3 CV, green line). (B) Gradient SDS-PAGE in for the fractions indicated on (A), the upper arrow marks the expected position for reduced protein (R) and the lower for non-reduced (NR).

Fractions 15-21 were pooled (final volume of 32 ml) and treated with Endo H_f and applied to the Con-A column as before. The flow-through from the Con-A column was concentrated to 10 ml and then loaded onto the size-exclusion chromatography column with injections ranging between 0.5 and 0.8 ml. Figure 5.36 (A) shows six of the injections overlaid.

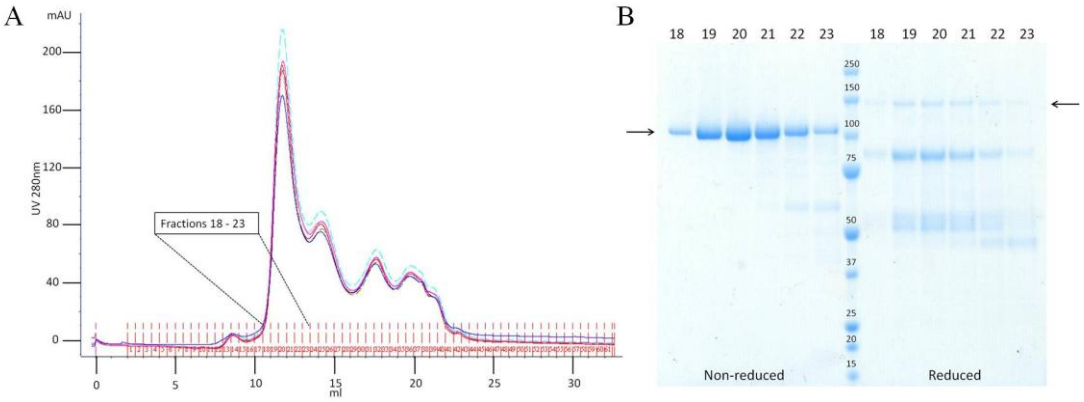


Figure 5.36 Size-exclusion chromatography of FH 3xGLY using a Superdex 200 HR 10/30 column. (A) The overlaid profiles; (B) The gradient SDS-PAGE for the fractions indicated on (A).

As expected, multiple peaks were obtained. The first major peak eluting between 10 and 13 ml corresponds to FH 3xGLY according to SDS-PAGE with a single band obtained under non-reducing conditions but the reducing conditions confirmed that like FH SL, FH 3xGLY also suffered from a large amount of clipping that yielded 80 and 60-KD fragments. Fractions 18-21 were pooled and concentrated to yield a final concentration of 4.6 µM in 1.5 ml.

5.4 Biophysical analysis

The ^{15}N HSQC for the sample of FH 12-13 WT (Figure 5.37) produced, in the current study, from codon-optimized DNA was initially compared to the assigned ^{15}N HSQC for the previously prepared FH 12-13 sample that had been used for structure determination [55]. The two spectra were almost identical, with slight differences probably due to small discrepancies in pH or salt (Figure 5.38).

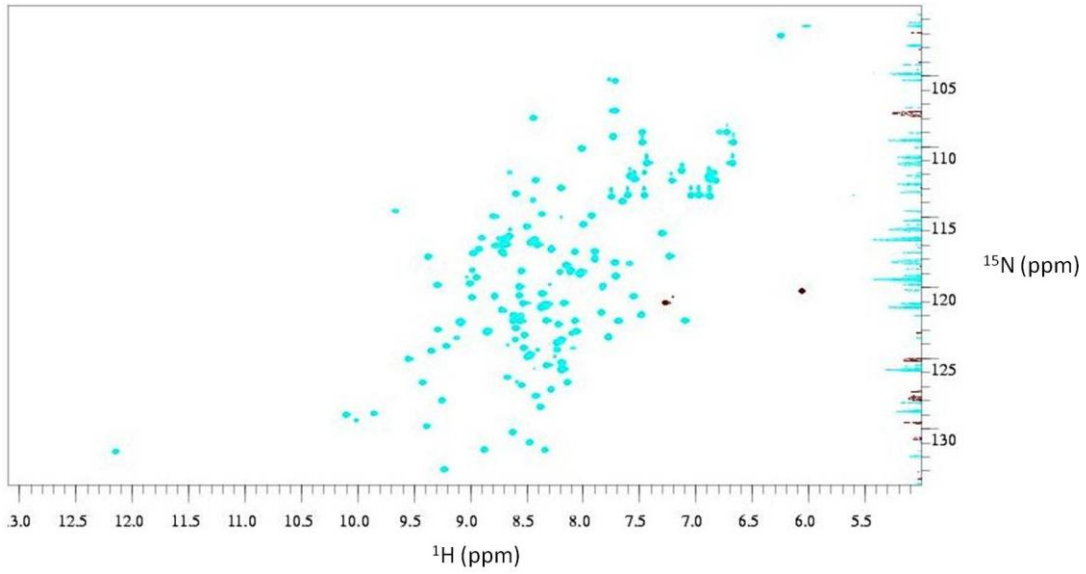


Figure 5.37 ^{15}N HSQC of FH 12-13 WT. Spectrum shows good dispersion with well resolved peaks. NMR sample is 200 µM in 20 mM potassium phosphate buffer pH 6.6. The purple peaks are spectrally folded.

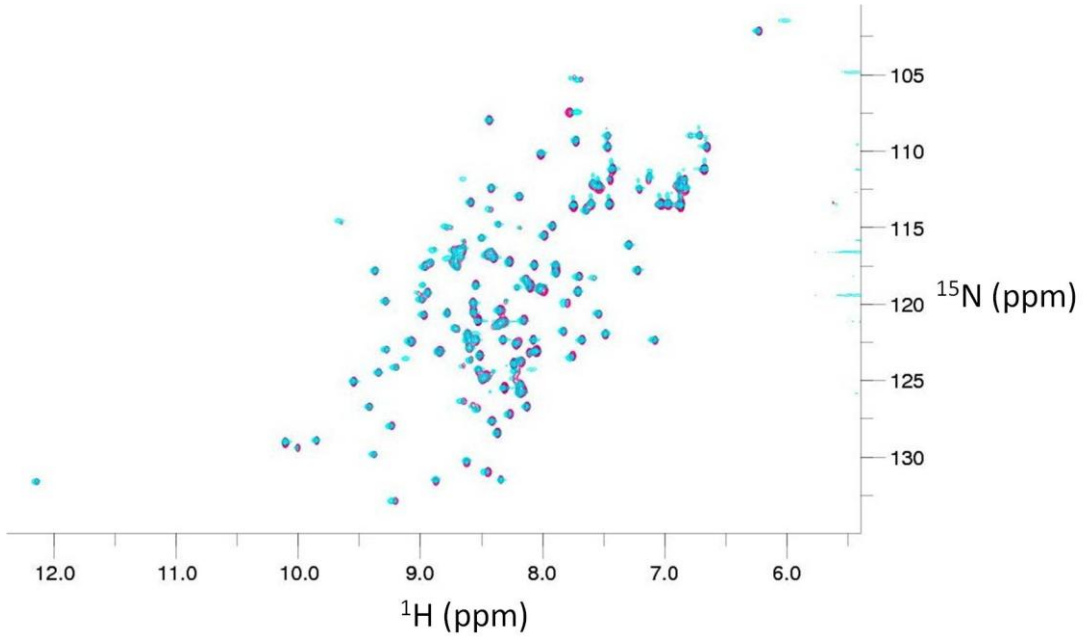


Figure 5.38 ^{15}N HSQC of the FH 12-13 WT sample prepared in the current work overlaid on the ^{15}N HSQC of FH 12-13 sample prepared previously. New FH 12-13 sample is shown in cyan; previous FH 12-13 sample shown in pink.

In the case of FH 12-13 3xGLY, the ^{15}N HSQC (Figure 5.39) again showed well-defined peaks indicating both of the CCPs were properly folded. When the spectrum was overlaid with the original WT data (Figure 5.40) there were again small shifts

observed, presumably due to pH or salt differences, but there are also bigger difference visible for some peaks, suggesting that the environment around these nuclei was different.

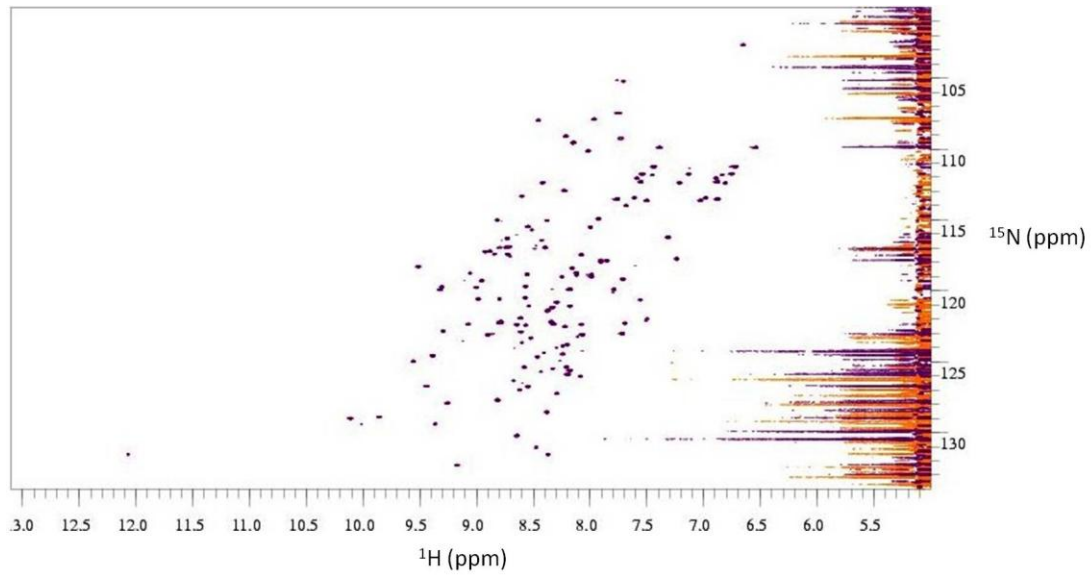


Figure 5.39 ^{15}N HSQC of FH 12-13 3xGLY. NMR sample is ~50 µM in 20 mM potassium phosphate buffer pH 6.6.

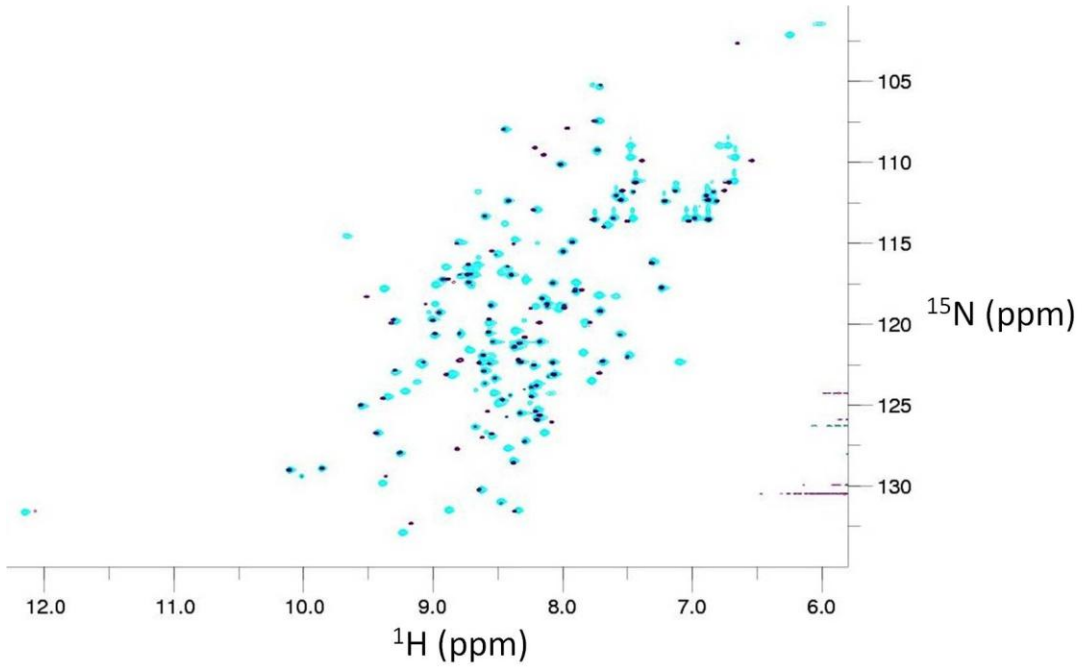


Figure 5.40 Overlaid ^{15}N HSQC spectra of FH 12-13 3xGLY and FH 12-13 WT. 12-13 3xGLY is shown in purple; FH 12-13 WT in cyan.

The ^{15}N HSQC for FH 12-13 SL (Figure 5.41) also showed well-dispersed and well-resolved cross-peaks, consistent with properly folded modules. When overlaid with the equivalent spectrum of FH 12-13 WT (Figure 5.42), the majority of peaks coincided well, but there were again clear differences for some peaks.

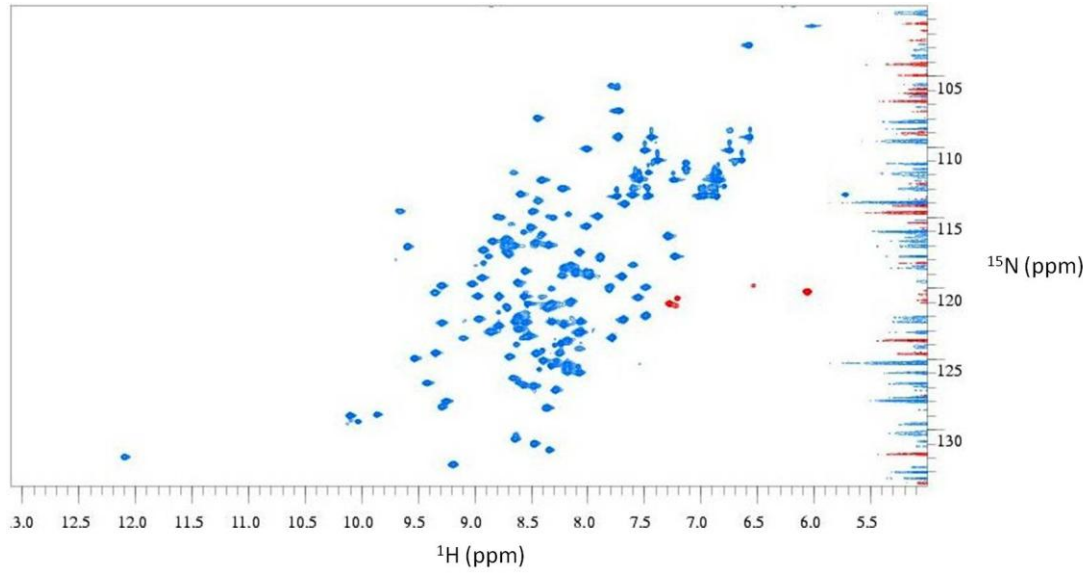


Figure 5.41 ^{15}N HSQC of codon optimized FH 12-13 SL. Spectrum shows good dispersion with well resolved peaks. NMR sample is 200 µM in 20 mM potassium phosphate buffer pH 6.6. The red peaks are spectrally folded.

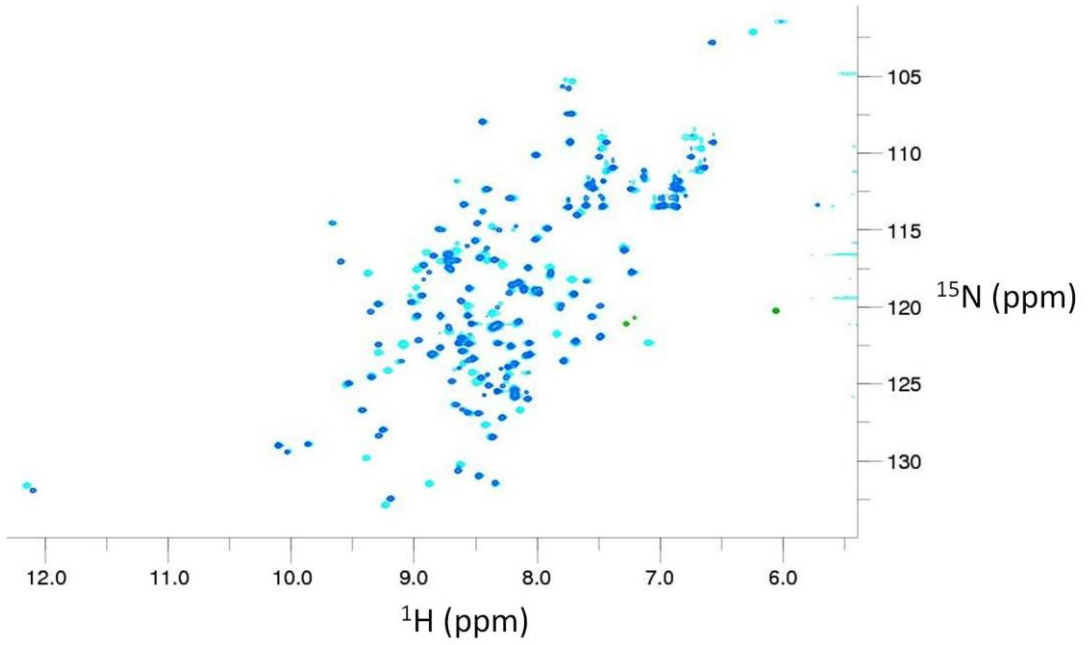


Figure 5.42 ^{15}N HSQC of FH 12-13 SL overlaid on the ^{15}N HSQC of FH 12-13 WT. 12-13 SL is shown in blue and green; FH 12-13 WT in cyan.

The differences between spectra for mutant and wild-type FH 12-13 were anticipated to coincide with residues in or close to the intermodular linker. Linker residues Val_{745}, Lys_{751} and Lys_{752} had not been mutated. As expected, no cross-peaks could be detected (in the spectrum of either mutant) for these three residues at their chemical shift, as previously assigned in FH 12-13 WT. Likewise, no cross-peaks appeared in the 3xGLY spectrum for remaining linker residues Ala746 or Asp748 at their previously assigned chemical shifts (Figure 5.43).

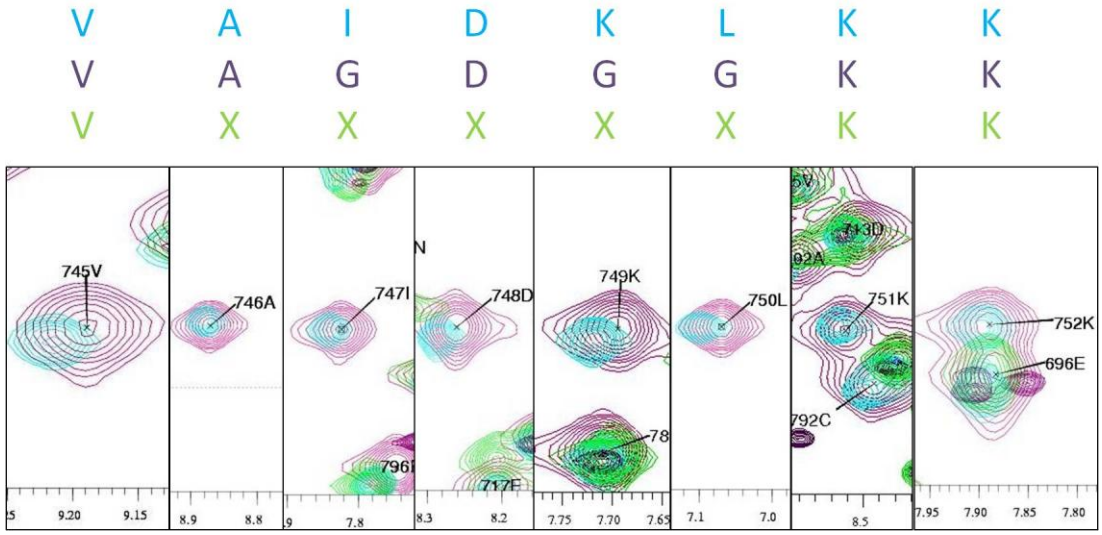


Figure 5.43 HSQC cross-peaks for linker residues. The pink and cyan peaks (and cyan one-letter code) correspond to the original and new WT samples, respectively. Purple peaks and one-letter code correspond to 3xGLY, while lime green peaks and one-letter code correspond to SL. X means these residues have been deleted.

5.4.1 Chemical shift perturbation of FH 12-13 3xGLY and SL

In order to investigate further if the structural changes caused by modifying the intermodular linker are confined only to residues within the interface between the CCPs chemical shift perturbations were calculated. The caluculations were performed using the ^{15}N-HSQC of FH 12-13 WT alongside either the ^{15}N-HSQC of FH 12-13 3xGLY or SL and the "shift differences" function in CCPN Analysis. Chemical shift values above the baseline ($\Delta\delta > 0.04$ ppm) were represented as putty models based on the NMR solution structure of FH 12-13 (PBD: 2KMS).

For FH 12-13 3xGLY (Figure 5.44) the largest changes in chemical shifts are only located within the interface between the CCPs (CCP 12 residues 720-727, CCP 13

residue 792-797). Again for FH 12-13 SL (Figure 5.45) the largest differences are within in the interface (CCP 12 residues 724-727, CCP 13 residue 793-797) but there are also some minor differences within CCP 12. In order to understand further the effects the linker modifications have on the structure of FH 12-13 a double labeled sample (^{13}C, ^{1}H) of 3xGLY and SL would be required which would allow the linker residues to be included and thus fully calculate any changes that are specific to the difference in the linker region.

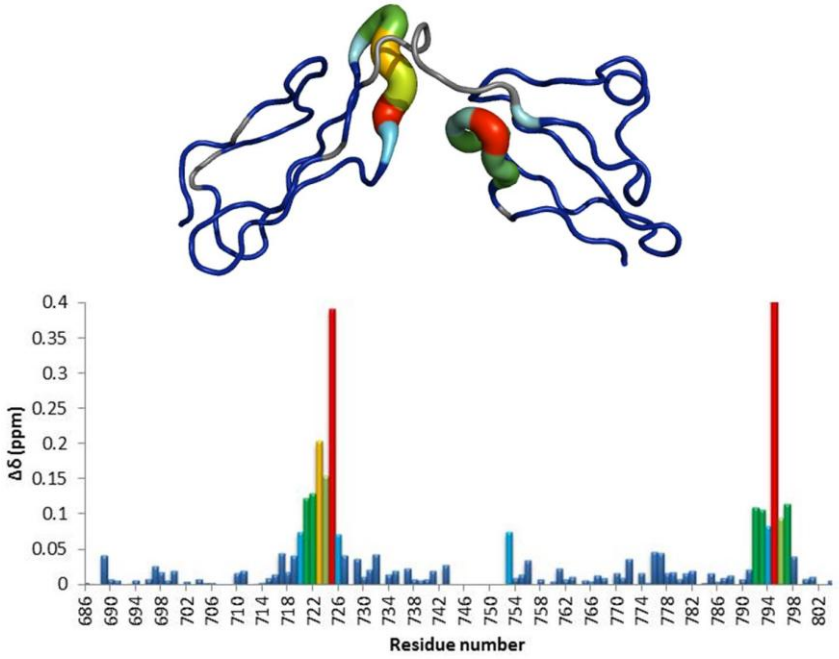


Figure 5.44 ^{1}H, ^{15}N Chemical shift differences for FH 12-13 3xGLY. The combined chemical shift differences for each residue are plotted as a function of residue number. The largest shift changes (Δδ > 0.04 ppm) are mapped onto the FH 12-13. The bale blue peaks correspond to residues between 0.05 and 0.08 ppm, green for 0.08-0.12 ppm, yellow for 0.12-0.20 ppm and red for above 0.2 ppm. Grey residues are either prolines or linker residues which there is no data for.

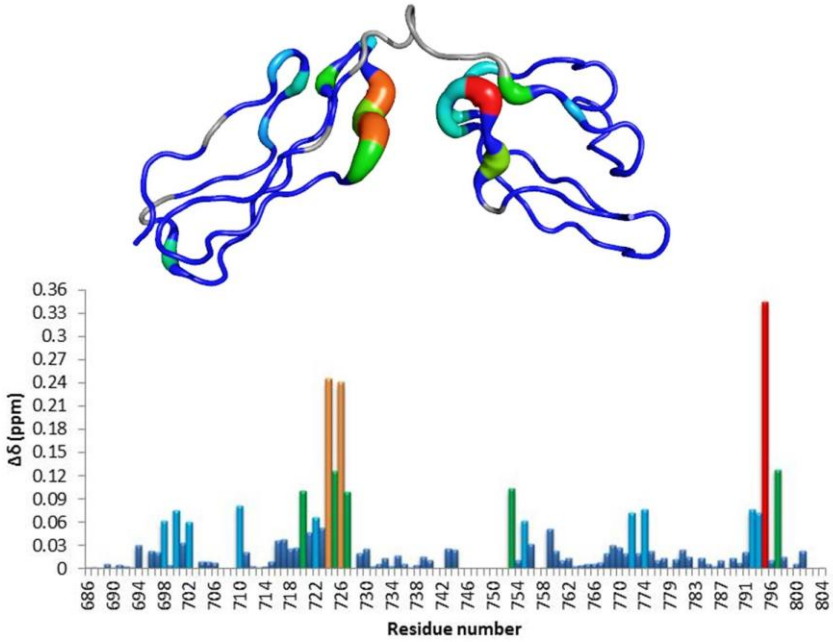


Figure 5.45 ^{1}H, ^{15}N Chemical shift differences for FH 12-13 SL. As above but the orange peaks correspond to shifts between 0.12 and 0.24.

5.4.2 Flexibility of FH 12-13 WT, 3xGLY and SL

Previous work on the flexibility within FH 12-13 revealed a relatively rigid structure [55] with little evidence for intermodular motion. The removal of bulky side-chains in 3xGLY was expected to introduce substantially more flexibility. On the other hand, the trimming of the linker to three residues (SL) would be expected to disrupt the intermodular interface, disallow the folded back structure of FH 12-13 WT but restrain flexure between modules.

For each construct, T_1, T_2 and ^{1}H, ^{15}N (heteronuclear) NOEs were recorded (Table 5.2 and Figures 5.46 – 5.48). Note that for FH 12-13 3xGLY, the data quality was poor for the T_1 and NOEs although the T_2 data were of adequate quality.

Table 5.2 Mean values of the relaxation times for all three FH 12-13 constructs.

		Mean T_1 (±SD) [ms]	Mean T_2 (±SD) [ms]	Mean NOE (±SD)
WT	Both CCPs	620 ± 79	76 ± 13	0.74 ± 0.11
	CCP 12 alone	630 ± 68	75 ± 7	0.75 ± 0.14
	CCP 13 alone	601 ± 92	75 ± 10	0.72 ± 0.05
SL	Both CCPs	658 ± 93	106 ± 25	0.64 ± 0.12
	CCP 12 alone	700 ± 86	99 ± 21	0.56 ± 0.07
	CCP 13 alone	606 ±70	112 ± 25	0.71 ± 0.11
3xGLY	Both CCPs	-	127 ±12	-
	CCP 12 alone	-	122 ± 25	-
	CCP 13 alone	-	130 ± 31	-

For each data set, the mean value and standard deviation (SD) was determined for the individual CCPs as well as the pair.

The mean values for T_1, T_2 and heteronuclear NOE for FH 12-13 WT are similar for both CCPs 12 and 13. This means there is no evidence for the modules moving independently, which is consistent with their being fused together. Similar results were obtained previously for FH 12-13 [55] and the NMR-derived solution structure supported the conclusion that FH 12-13 has a relatively rigid structure [55].

In the case of FH 12-13 SL, relaxation values for the two modules appear to differ suggesting that the CCPs move independently of one another. Thus despite having only a three-residue linker, this mutant is more flexible presumably because – as anticipated - there is no longer an extensive set of intermodular and module-linker interactions present in the structure. The lower NOE values for CCP 12 could also reflect loss of rigidity within this module that is normally stabilized by its interface with CCP 13. Looking at the distribution of decreased heteronuclear NOE values in the structure does not really support this hypothesis since they are not concentrated near the interface but a fuller assignment would be needed to draw firm conclusions.

Only T_2 data were available for FH 12-13 3xGLY. The average values for the CCPs 12 and 13 appear to differ in this mutant, so this could hint at more intermodular flexibility; but this would need to be confirmed by preparing another sample to re-

record the T_1 and heteronuclear NOE experiments. Note that FH 12-13 3xGLY has significantly higher T_2 values than does FH 12-13 wt. Overall, the T_2 values suggest that FH 12-13 3xGLY has the most intermodular, and/or intramodular, flexibility, followed by FH 12-13 SL, with FH 12-13 WT being the least flexible.

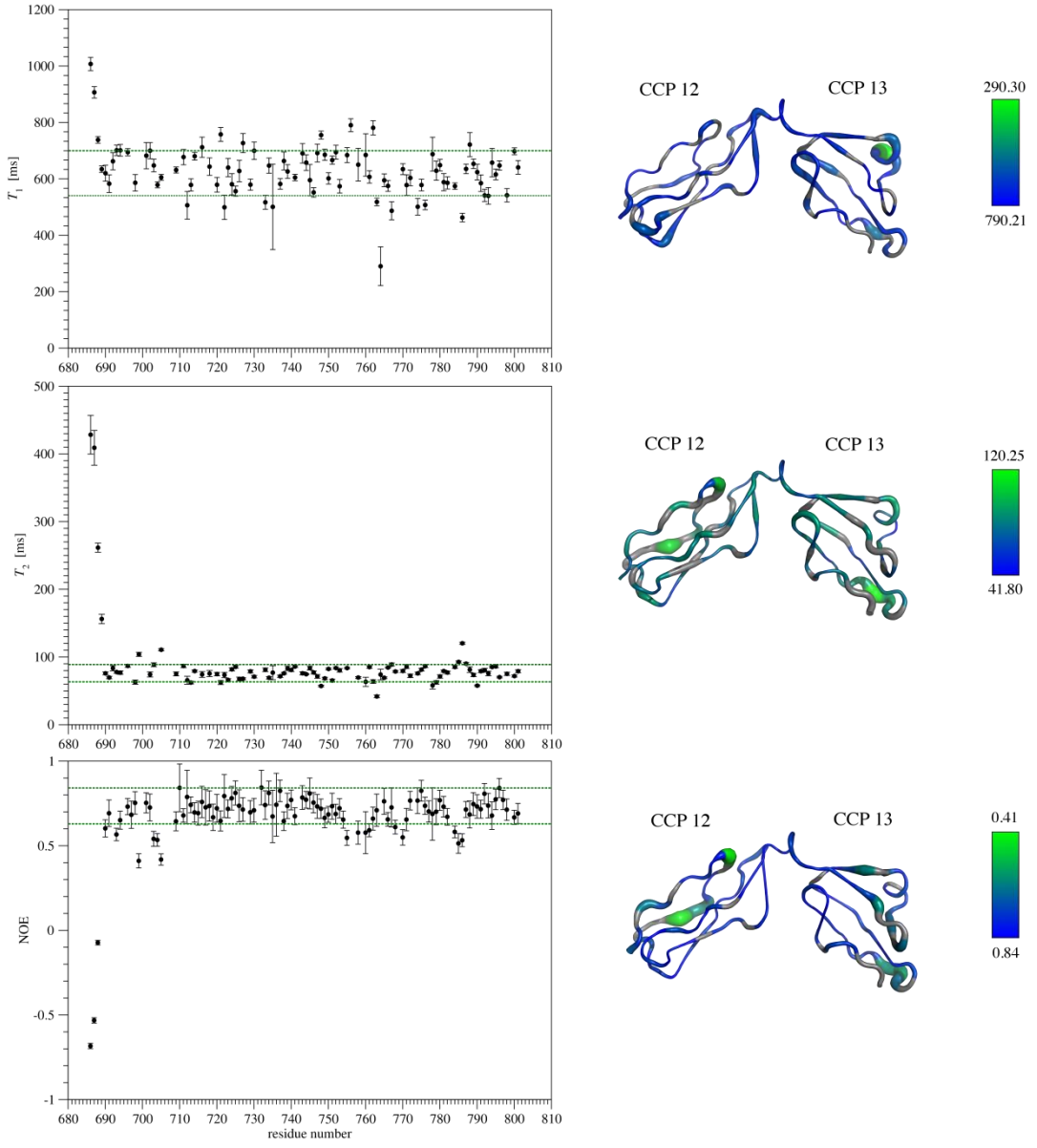


Figure 5.46 12-13 WT, T_1, T_2 and NOE values. The broken green lines indicate one SD either side of the mean. The corresponding putty diagrams are based on PDB model 2KMS. Thickness and color of the backbone traces corresponds to the putative flexibility suggested by the data. Green signifies more flexible, blue indicates more rigid. Grey indicates residues for which data is unavailable due to spectral overlap, the presences of prolines, or very high peak broadening leading. Figures prepared with the help of Dr Mateusz Maciejewski.

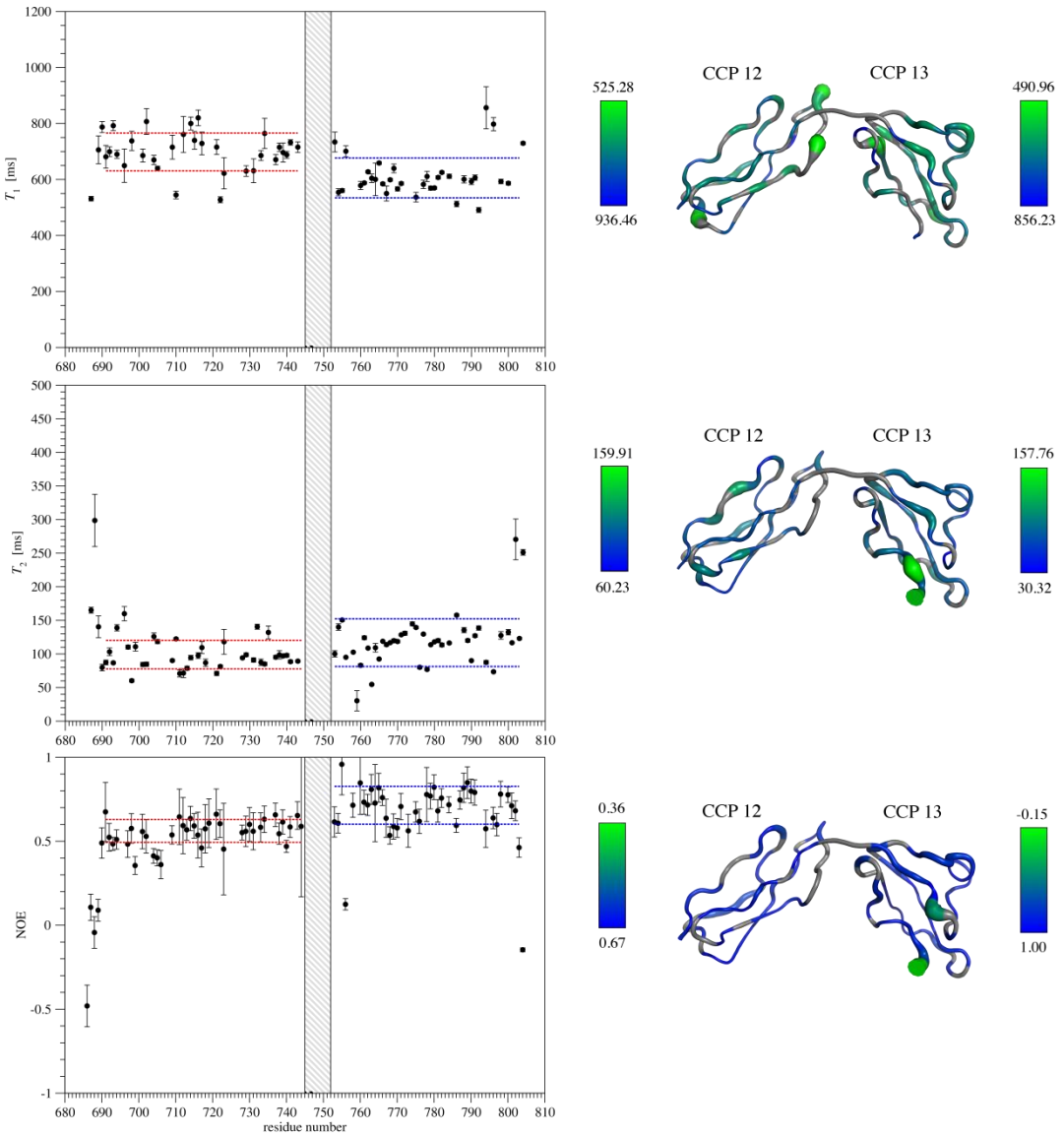


Figure 5.47 T_1, T_2 and NOE values for 12-13 SL. Red and blue dashed lines indicate one 1 SD on either side of the means for CCPs 12 and 13 respectively. The grey box in each plot represents the linker region, which could not be assigned due to the sequence difference with respect to WT. The residue numbers match those of WT in order to allow comparisons to other constructs. The putty models shown are based on a model of the SL variant constructed using Modeller v9.10 using PDB model 2KMS. The color coding of the models is identical to that used for FH WT in Fig. 5.44. Figures prepared with the help of Dr Mateusz Maciejewski.

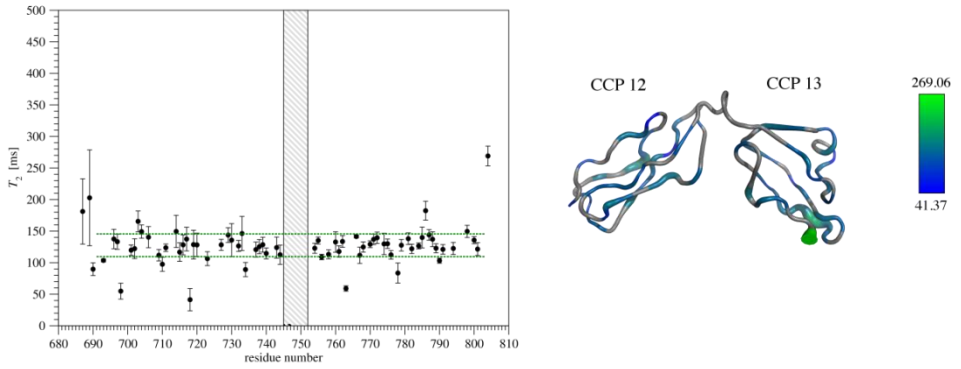


Figure 5.48 12-13 3xGLY T_2 values. The dashed green line limits the range of the T2 values within 1 SD of the mean. The shaded region represents the linker region which could not be assigned. The putty model was constructed using PDB 2KMS as the base via Modeller v 9.10. The color coding of the model is the same as the WT construct. Figure prepared by Dr Mateusz Maciejewski.

Diffusion tensor for WT and SL

In order to determine the diffusion tensor of a protein, the minimum data required are the T_1 and T_2 values; as the T_1s are not available for FH 12-13 3xGLY, this protein will not be considered any further.

Initially T_1/T_2 ratios were used to approximate the elements of the diffusion tensor for both FH 12-13 WT and FH 12-13 SL [177]. The ratios suggest that the diffusion tensors are either asymmetric or fully anisotropic because they are spread over a range of values (Figure 5.49).

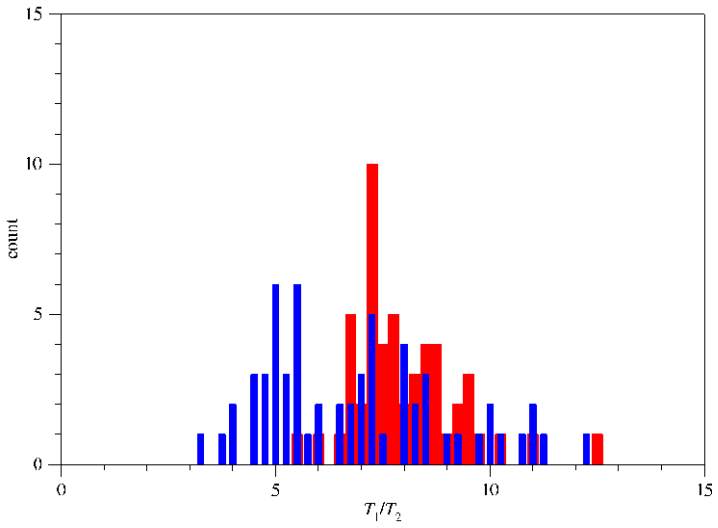


Figure 5.49 T_1/T_2 ratios for FH 12-13 WT and FH 12-13 SL. Blue bars are for SL and red for WT. (bin size for both was 0.25). Figure prepared with help of Dr Mateusz Maciejewski.

The data were then fitted to three different diffusion tensor models [178] to determine which was best. Models used were: the fully isotropic tensor, which is characterized using just the overall rotational correlation time (τ_c) (even though this looks unlikely as alluded to above); the axially symmetric tensor which requires four parameters - the overall rotational correlation time (τ_c), the tensor anisotropy ($D_{\parallel}/D_{\perp}$) and the Euler angles θ and ϕ; and finally the fully anisotropy tensor characterized by six parameters - the overall rotational correlation time (τ_c), the tensor anisotropy, the tensor asymmetry (D_{yy}/D_{xx}) and the Euler angles ψ, θ and ϕ. Throughout the fitting for FH 12-13 WT, the closest-to-mean structure of PDB:2KMS was used, and for SL structure 9 was used as it gave the lowest *E/N* values of all the models present in 2KMS. The results for each of the tensor models are shown in Table 5.3 for FH 12-13 WT and Table 5.4 for FH 12-13 SL.

Table 5.3 WT diffusion tensor fitting.

	Isotropic	Axially symmetric	Anisotropic
τ_c	8.46 ± 0.11	8.68 ± 0.03	8.60 ± 0.02
Anisotropy	1	1.46 ± 0.01	1.41 ± 0.03
Asymmetry	1	1	1.21 ± 0.01
θ (°)	-	111.40 ± 0.11	108.79 ± 0.16
ϕ (°)	-	177.85 ± 0.10	176.40 ± 0.19
ψ (°)	-	-	136.05 ± 0.26
F statistic	-	7.61	1.66
p value	-	2.9×10^{-4}	0.20

Table 5.4 SL diffusion tensor fitting.

	Isotropic	Axially symmetric	Anisotropic
τ_c	7.23 ± 0.15	7.23 ± 0.01	7.20 ± 0.02
Anisotropy	1	1.80 ± 0.01	1.71 ± 0.01
Asymmetry	1	1	1.39 ± 0.01
θ (°)	-	118.10 ± 0.02	120.70 ± 0.04
ϕ (°)	-	227.45 ± 0.01	225.61 ± 0.03
ψ (°)	-	-	261.53 ± 0.01
F statistic	-	4.97	1.93
p value	-	0.004	0.15

The F statistic and rejection probability (*p* value) for both WT and SL indicate that the axially symmetric tensor was preferred over the anisotropic and isotropic tensors. However for SL even though the axially symmetric model was preferred the fitting of the data to this model resulted in a very large *E/N* score of 94.6 (*N* is the number of residues used for the calculation, E is the total error) in comparison to the WT score of 7.8. As only the linker differs between the two it was thought that the high *E/N* score was due to an alteration in interdomain orientation. In order to improve this issue the data for each CCP was fitted separately. Whilst this improved the score for CCP 12 (6.5), the value for CCP 13 remained too high (47.4). There are two possible explanations for why this might be occurring: the first is that the CCPs of FH 12-13 SL do not – in fact - diffuse independently of each other and the other is that the CCPs in FH 12-13 SL diffuse together but with some interdomain flexibility.

To investigate if the second explanation could be true the diffusion tensor required some refinement. The parameters for the axially symmetric tensor were used (see

above) along with the slow-motion order parameter (S_S^2) and the slow-motion correlation time (τ_s) for each CCP – these parameters describe interdomain motion. The calculation was also supplemented with averages of the fast internal motion order parameter (S_f^2) and the fast internal motion correlation time (τ_f) averaged for all residues in both CCPs. A refinement of the WT diffusion tensor was also performed though it did not include parameters for interdomain motion (τ_s and S_S^2). Table 5.5 summarizes the results for the calculations.

Table 5.5 Refined tensor parameters for WT and SL.

	WT	SL
τ_c	8.40 ± 0.07	10.01 ± 0.54
Anisotropy	2.09 ± 0.07	1.76 ± 0.11
θ (°)	117.13 ± 1.72	128.96 ± 13.82
ϕ (°)	174.73 ± 2.82	232.99 ± 13.29
S_f^2	0.95 ± 0.01	0.085 ± 0.02
τ_f (ps)	184 ± 7	25 ± 15
S_S^2 [12]	-	0.72 ± 0.05
τ_s [12] (ns)	-	1.54 ± 0.26
S_S^2 [13]	-	0.65 ± 0.05
τ_s [13] (ns)	-	3.31 ± 0.51
E/N	0.23	0.57

Table 5.5 Refined tensor parameters for WT and SL. 12 and 13 inside square brackets represent CCP 12 or CCP 13.

The result of the refinement process for SL has greatly improved the E/N score to 0.57 (initially was 94.6) and the WT score has also improved (0.23 compared to 7.8 initially). The new parameters recorded for SL indicate a large range of interdomain mobility and this is supported by the slow motion order parameters of 0.72 and 0.65 which correspond to semi-cone angles of 39° and 44° respectively. The ranges are relative to the mean orientation in the diffusion frame and thus should be treated as the maximal ranges of tilt of the CCPs. Thus these data support our hypothesis that

shortening the linker by deleting residues with bulky side chains prevents the 80-degree tilt seen in FH 12-13 WT and at the same time disrupts intermodular rigidity, resulting in a more elongated and flexible interdomain arrangement.

Model-free and extended model-free analysis of relaxation data for WT and SL

The diffusions tensors in Table 5.5 were used to calculate Model free (MF) and extended model free (EMF) parameters for WT and SL. An approach similar to that assumed in the Modelfree software was used. In brief, the data for each residue was fitted to five different models. Model 1 was characterized by S_f^2, model 2 by S_f^2 and τ_f. Model 3 was S_f^2 and the chemical exchange rate R_{ex}, for model 4 it was S_f^2, τ_f and R_{ex}. Finally model 5 was characterized by S_f^2, τ_f and S_S^2. For SL only models 1-3 will be considered due to the slow motion parameters already being included in the diffusion tensor for SL.

Model 1 was used when the *E* value from the fitting of the relaxation data to this model was within the 10% confidence interval; models 2 or 3 were used if *E* from the fittings was again within the 10% confidence range but also was shown by an *F*-test to fit the data better than model 1. Models 4 and 5 were considered if the *E* value was larger than the 10% confidence interval. As the number of experiments for each residue was 3 (T_1, T_2 and NOE) and the number of parameters in models 4 and 5 was also 3 the number of degrees of freedom was set at 0. It was therefore expected that the E values would be approximately 0 but if this was not true for both models 4 and 5 then it was concluded that no model provided an appropriate description of the dynamics for that given residue.

The results for the dynamic parameters for WT and SL are shown in Figures 5.50 and 5.51. The large errors seen in plots of τ_f and R_{ex} for WT show that the values only reflect the rough order of the timescale of these parameters for the relevant residue. In the case of SL, model 1 was preferred for all residues.

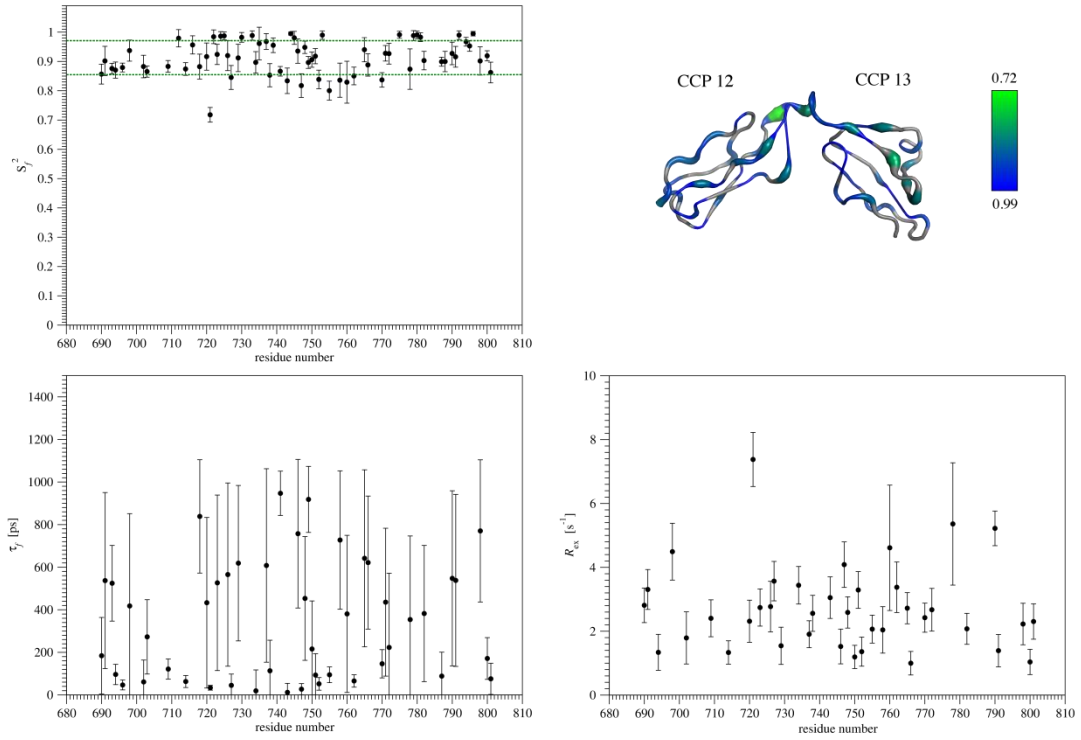


Figure 5.50 MF parameters for WT. The error bars designate the distance one SD away from the mean.

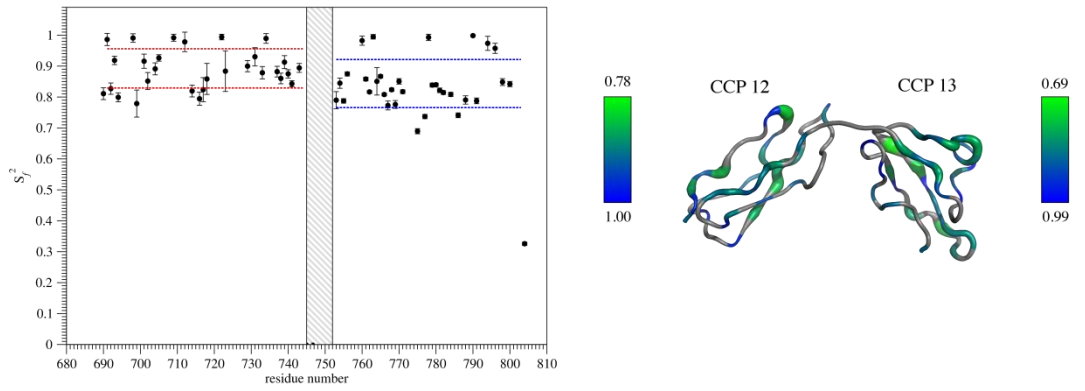


Figure 5.51 MF parameters for SL. The error bars designate the distance one SD away from the mean. The shaded box represents the linker region.

In conclusion the results for the analysis of the interdomain flexibility within the three FH 12-13 constructs confirmed that WT did indeed have limited flexibility as previously observed. The results for the SL mutant suggest that there is movement between the domains.

5.4.2 SAXS analysis of 8-15 and 10-15 linker constructs

5.4.2.1 Introduction

SAXS is an established technique for the analysis of biological macromolecules in solution. A typical biological experiment involve the macromolecules being randomly order in a solution which are then illuminated by a monochromatic and collimated X-ray beam, the resulting scattering pattern is then recorded by an X-ray detector. The scattered intensity I is recorded by the dector as a function of the scattering angle, 2θ (Figure 5.52).Scattering from the solvent (e.g. dialysis buffer) is also recorded and then subtracted from the measured sample. The resulting scattering profile is generally expressed as a function of the momentum transfer ($s=4\pi\sin\theta/\lambda$, λ is wavelength) and is proportional to the average scattering from a single molecule in all available orientations as well as to the concentration of the solute.

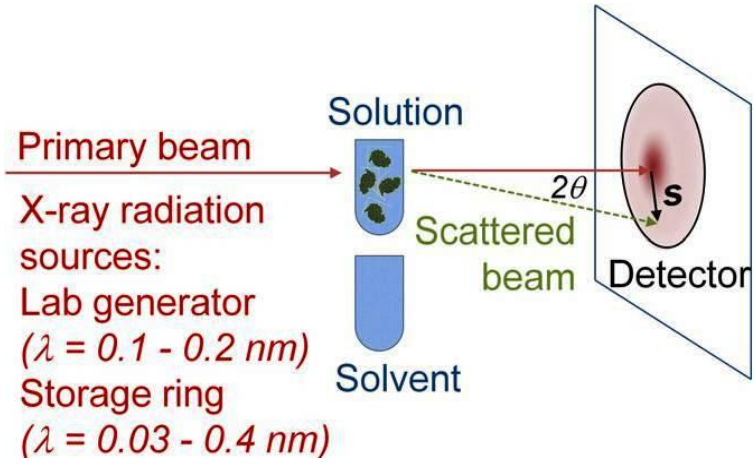


Figure 5.52 Schematic representation of a biological solution SAXS experiment. Image from [179]

Information that can be extracted from the scattering profile includes molecular weight (MWt), radius of gyration (R_g), hydrated particle volume (V) and the maximum particle diameter (Dmax) [179, 180]. The experiments required to obtain this information are highlighted in Table 5.6

Table 5.6 Details of SAXS experiments used.

SAXS Experiment	Characteristics of sample
Guinier plot (calculate R_g and $I(O)$)	Purity
Pairwise interatomic distance distributions, $p(r)$	Changes in shape (displayed graphically)
Kratky plot	Flexibility (initial look)
Multiple conformations (EOM)	Flexibility
A*b initio* models	Shape (displayed using models)
Overall parameters (R_g, D_{max} and MWt)	Validation of structure

Prior to sending the samples for SAXS analysis, they were checked for the presence of oligomers or aggregates using dynamic light scattering. According to the size-distribution by intensity analysis for the FH 8-15 mutants, the 3xGLY and SL protein samples appeared to be monodisperse but the native-sequence (WT) protein contained very minor amounts of aggregates (although these were not detected in the less sensitive size-distribution by volume analysis). The results for the FH 10-15 proteins, for both types of analysis, showed that no oligomeric impurities were detectable (Figure 5.53).

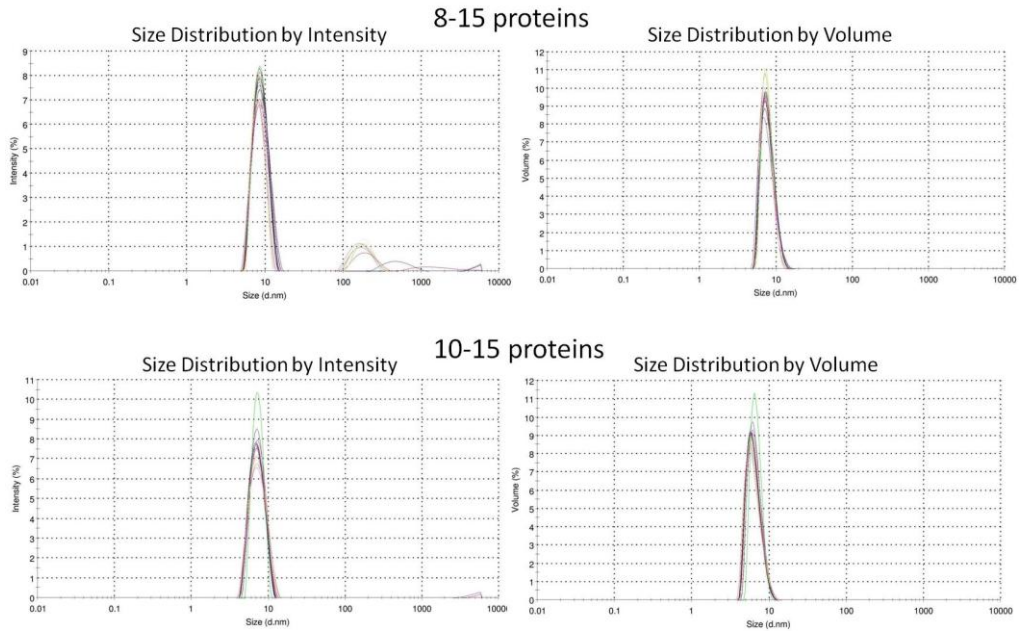


Figure 5.53 DLS data for FH 8-15 and FH 10-15 protein constructs. The top plots are for the 8-15 proteins and the lower ones are for 10-15 proteins.

5.4.2.2 SAXS results for FH 8-15 and FH 10-15 linker constructs

The scattering profiles obtained for all the constructs are shown in Figure 5.54 (FH 8-15) and Figure 5.55 (FH 10-15). Both sets of curves yielded a linear Guinier plot for each construct, which is a good indication that the samples are of good quality and therefore a range of parameters could be legitimately extracted (Table 5.7).

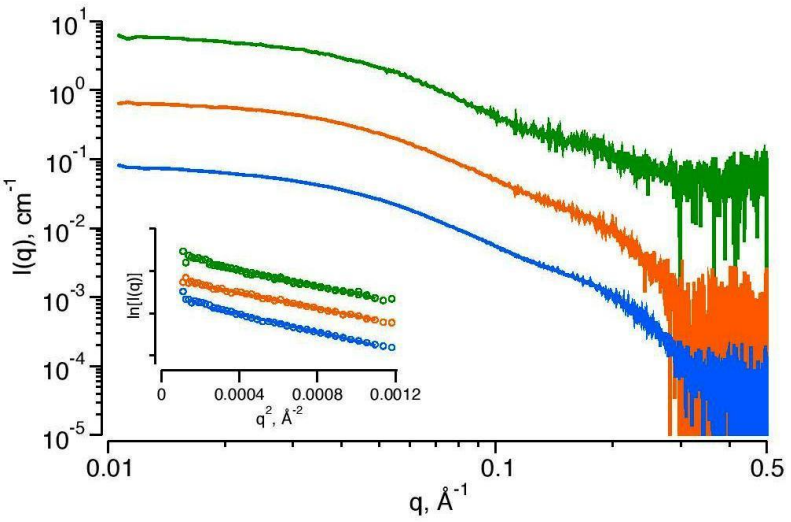


Figure 5.54 SAXS profile for FH 8-15 linker constructs. The green scattering curve represents WT, orange represents 3xGLY and blue represents SL. Inserted are the corresponding linear Guinier plots. The y-axis has been adjusted to show all three plots in a non-overlapping way for display.

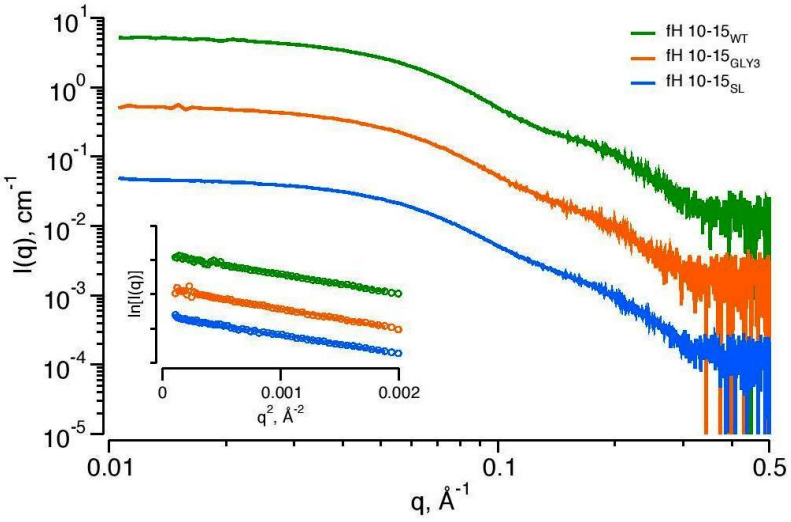


Figure 5.55 SAXS scattering curves for FH 10-15 linker constructs with linear Guinier plot inserted. See Figure 5.53 for legend.

Table 5.7 SAXS parameters of FH 8-15 and 10-15 linker constructs.

FH Construct	*Rg*, A	*I(0)*, cm^{-1}	D_{max}, A	*Vp* ($x10^3$), A^3	*MWt*, kDa[a]
10-15 WT (41.1 kDa)	29.3 ± 0.1	0.05 ± 0.01	103 ± 5	68 ± 10	42.5 ± 5
10-15 3xGLY (41.1 kDa)	30.3 ± 0.3	0.04 ± 0.01	106 ± 5	69 ± 10	43.5 ± 5
10-15 SL (41.1 kDa)	28.5 ± 0.1	0.05 ± 0.01	98 ± 5	59 ± 5	36.8 ± 5
8-15 WT (54.7 kDa)	36.2 ± 0.1	0.06 ± 0.01	127 ± 5	115.2 ± 10	74.5 ± 10
8-15 3xGLY (54.7 kDa)	42.6 ± 0.1	0.09 ± 0.01	136.4 ± 5	131.8 ± 10	82.4 ± 10
8-15 SL (54.7 kDa)	44.2 ± 0.1	0.10 ± 0.01	148 ± 5	147.5 ± 10	92.2 ± 10

Radius of gyration (*Rg*), forward scattering intensity (*I(0)*) determined from Guinier analysis. Maximum particle dimension (D_{max}) determined via indirect Fourier transformation. [a]Molecular weight (*MWt*) estimated based upon the hydrated particle volume (*Vp*).

The parameters extracted for FH 10-15 WT were compared against those already published in [55] and were found to be in good agreement. Analysis of the parameters extracted for all three FH 10-15 constructs suggested that there was little difference between the three and the distance distribution function p(r) for each (Figure 5.56) also showed only very small differences.

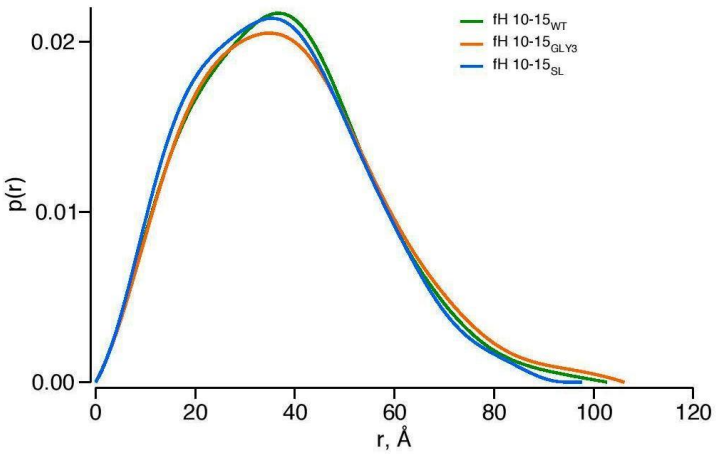


Figure 5.56 Pairwise interatomic distance distributions, *p(r)* for the FH 10-15 constructs.

As for the FH 10-15 constructs there were only very small differences between the FH 8-15 constructs apparent from SAXS analysis (Table 5.7 and Figure 5.57).

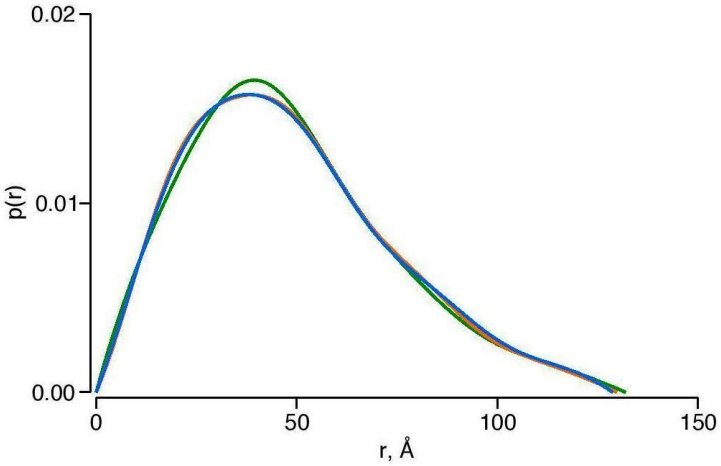


Figure 5.57 distance distribution function p(r) for FH 8-15 constructs. FH 8-15 WT is represented by the green line, for 3xGLY the orange line and SL is the blue line.

The results so far indicate that there are very few differences between the three constructs for both FH 10-15 and FH 8-15. The next step was to investigate if there was a change in flexibility between the constructs as had been seen in the NMR-based comparisons of the FH 12-13 proteins. First a Kratky plot was used to have an initial look at the flexibility within all the proteins. For FH 8-15 WT (green line, Figure 5.58) there is the peak from the low-angle scattering followed by a continuous increase as the scattering angle increases, which is the expected result for a multi domain protein [180]. On the other hand, with the 3xGLY (orange line) and SL (blue line) samples, as the scattering angle increases there is a decrease in scattering angle. This could be due to a poor match between the blank buffer and the sample buffer.

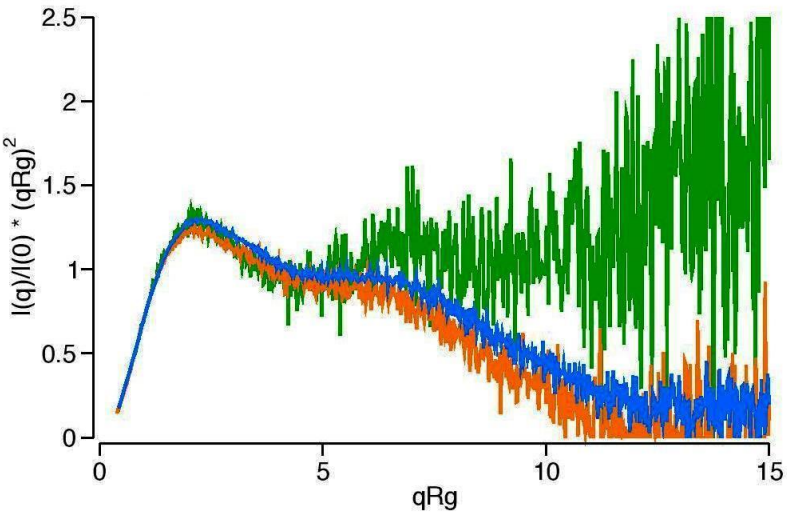


Figure 5.58 Kratky plot for FH 8-15 constructs. As before green is WT, orange 3xGLY and blue SL.

The Kratky plot results for FH 10-15 constructs (Figure 5.59) was much better as all three behaved as would be expected for a multi-domain protein.

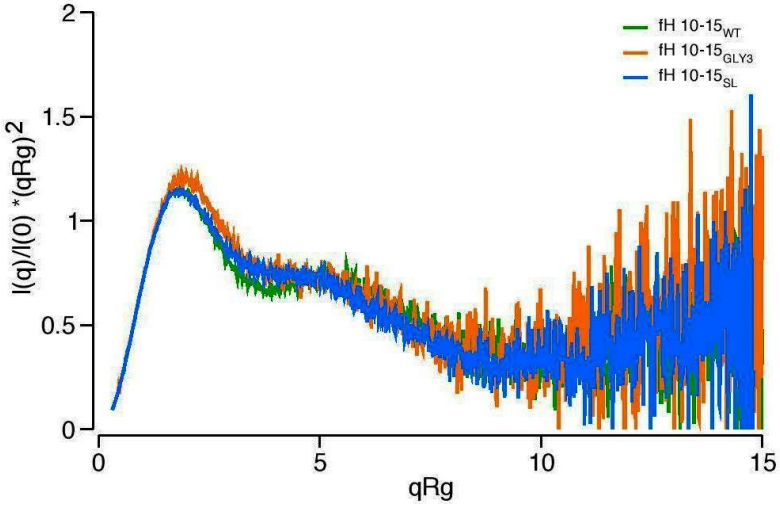


Figure 5.59 FH 10-15 constructs Kratky plot.

As the results for the FH 10-15 constructs were good it was possible to use the SAXS-based ensemble optimization method (EOM) in an attempt to investigate if there were any differences in flexibility within the three constructs. The method works by using a large pool of random configurations from which ensembles are selected using a genetic algorithm [181]. For all three constructs the data was

confined to the left side of the pool (Figure 5.60), this indicates that there is similarly limited flexibility within all three constructs. Thus the extra flexibility between CCPs 12 and 13 allowed by the mutated linker is not manifested in these longer constructs.

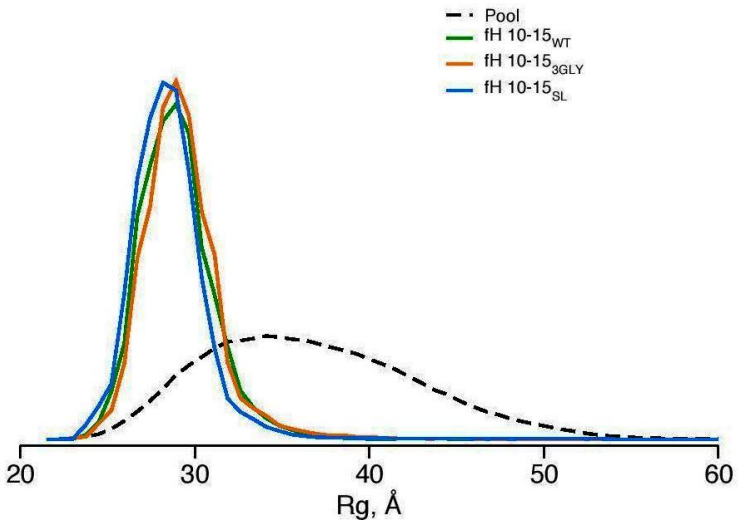


Figure 5.60 EOM analysis of FH 10-15 constructs.

As a final approach to see if there were any SAXS-detectable differences between the FH 10-15 constructs, an *ab initio* approach was used. The bead modeling program DAMMIF was employed to reconstruct an average structural model for each construct [182]. The results from this exercise (Figure 5.61) suggest that all three form similar compact structures. This is interesting since it means that the more extended conformation of CCPs 12-13 induced by the (engineered) short linker between these modules does not cause a loss of the compact overall arrangement of the longer construct. This implies that there are sufficient intermodular interactions within FH 10-15 – most likely between non-neighbouring modules - to accommodate the mutations and resist an opening out or "springing open" effect of the mutated linker.

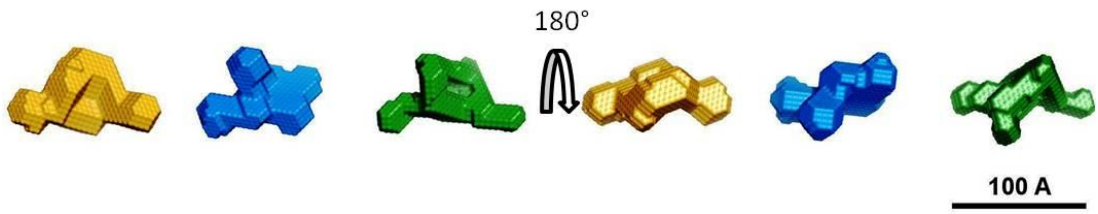


Figure 5.61 A*b initio* DAMMIF models for FH 10-15 constructs. The yellow models are for 3xGLY, blue for SL and green for WT.

5.5 Conclusions

The work described in this Chapter was performed to investigate the hypothesis that the long linker located between CCPs 12 and 13, which forms a "mini-domain" resulting in a relatively rigid and bent intermodular orientation, is important for FH architecture. An investigation of functional repercussions is reported in the next Chapter.

The initial work focused on FH 12-13 because detailed structural data were already available for the native-sequence (WT) construct and therefore any changes that resulted from mutagenesis of the linker sequence could be easily identified [55]. In this part of the work, ^{15}N HSQC spectra were recorded since chemical shifts allow small differences in structure to be inferred. The results showed, importantly, that radical engineering of the linker residues did not perturb the compact folding of the modules on either side. On the other hand, NMR-based ^{15}N-relaxation measurements showed significant differences in dynamic properties of the mutants. Although a complete data set for FH 12-13 3xGLY was not obtained, this mutant has much longer T_2 values than WT consistent with less overall rigidity. It would be fascinating to explore this observation further by recording more data on a new sample. The FH 12-13 SL mutant was also more dynamic and in this case a more rigorous analysis was possible. This implied that, consistently with intuition, removal of the residues forming the mini-domain between CCPs 12 and 13 resulted in a module pair that was less tilted and more flexible, leading to the expectation of a profound effect of such an engineering exercise on the architecture of the parent FH molecule.

As it is difficult to interpret SAXS data in terms of individual modules and module-module orientations for a protein as large as full-length FH, the same linker mutations were also made in the contexts of the shorter, more tractable constructs, FH 10-15 and FH 8-15, which encompass the central region of FH. According to SAXS, the FH 10-15 WT construct has a relatively rigid and compact structure (as observed previously) but, interestingly and unexpectedly, so did the FH 10-15 SL and FH 3xGLY constructs. As no real differences between FH WT, SL and 3xGLY were observed in terms of overall shape, it follows that interactions between non-contiguous CCPs must dominate these compact structures that can therby accommodate differences in one linker. Unfortunately the SAXS data for FH 8-15 were of lower quality due to poor buffer matching and degradation of this protein in the X-ray beam. Very recent SAXS data (not shown) with better buffer matching indicated that FH 8-15 SL (although still subject to degradation) had a significantly greater D_{max} than FH 8-15 WT. This is intriguing and not necessarily incompatible with our interpretation of the SAXS data for FH 10-15 WT *versu*s FH 10-15 SL. The resolution of SAXS is insufficient to uncover differences in the relative positions of specific modules within similar compact arrangements present within both FH 10-15 WT and FH 10-15 SL, but these might on the other hand be revealed by the additional presence of CCPs 8-9 that in FH 8-15 WT appeared to project away from the core formed by CCPs 10-15. These hypothesis could be tested by further SAXS data collection on fresh samples (of FH 8-15 but also on full-length FH) or could be tested by an orthogonal methods such as the use of chemical cross-linking combined with mass spectrometry to identify cross-linked peptides, as has proved successful recently (unpublished) with samples of intact FH.

In summary, engineering the CCPs 12-13 linker provides a route to versions of FH that have intact CCP modules and maintain a compact arrangement of their central modules but could have differences in flexibility and overall architecture. These make intriguing subjects for functional studies as explored in the next Chapter.

Chapter 6 Investigating the functional role of the central region of FH

The previous chapter described a study of the extent to which mutations of a linker within the central region of FH alter its flexibility and architecture. This chapter describes the functional analysis of the 3xGLY and SL mutants in the context of FH 8-15 and full-length FH.

Surface plasmon resonance (SPR) was used to measure the affinities of plasma-purified and recombinant FH (wild-type, plus the SL and 3xGLY mutants) for the following fragments of C3 immobilized on a sensor chip: the large (180 kDa), opsonic, activation-specific fragment, C3b, which binds covalently to surfaces; the soluble and seemingly inert product of iC3b cleavage (by factor I) C3c; and C3d - equivalent to the thioester domain (TED) of the original C3b - which is the final product of C3b cleavage, and remains tethered to the surface. In addition, an SPR-based assay was used to compare decay-accelerating activity, *i.e.* the enhancement of the rate of dissociation of C3b.Bb, between wild-type and mutant versions of FH.

Recent work in the lab had also created the possibility of performing binding studies for the recombinant FH 8-15 segment since this encompasses a binding site for the bacterial protein PspC. Furthermore, binding of PspC to central modules of full-length FH alters its overall architecture and boosts the regulatory activity of FH, providing further opportunities for functional characterization of the full-length FH mutants.

The proteins used in these studies are shown in Figure 6.1.

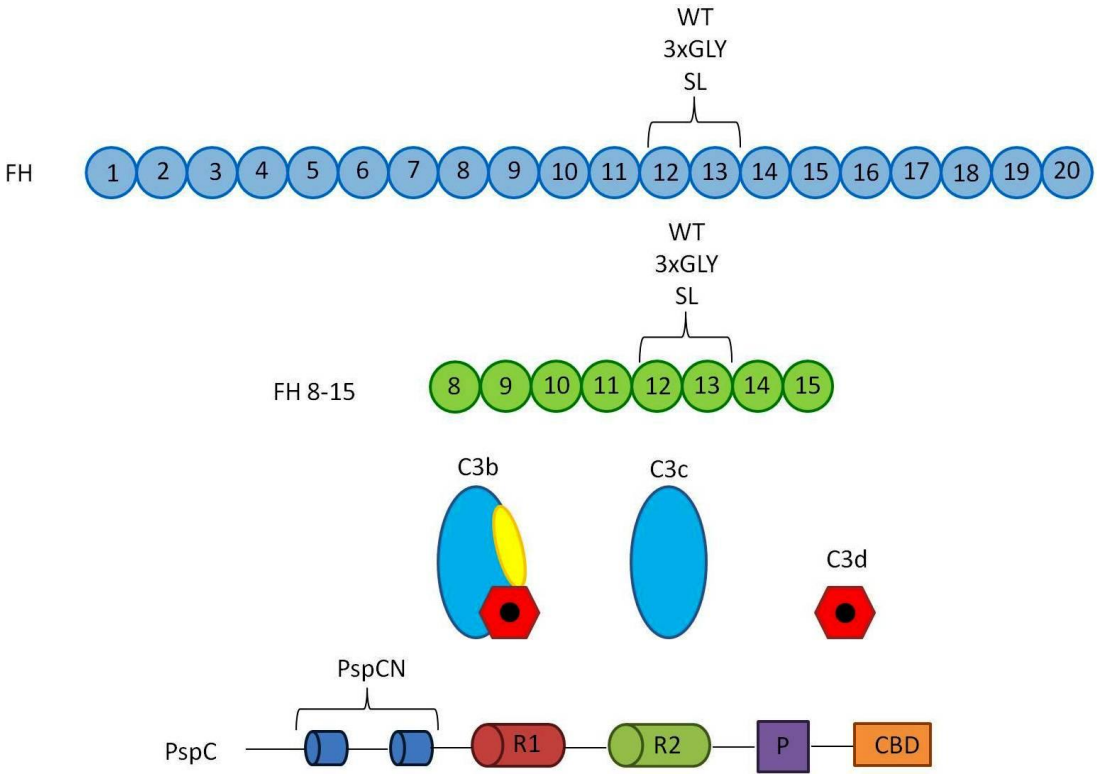


Figure 6.1 Summary of proteins used in the study. Full-length FH (FH) and FH 8-15 are shown with each oval representing a CCP module; the 12-13 linker modifications are shown above CCPs 12 and 13. Representation of the C3 fragments used; C3b, C3c and C3d. PspC is represented with the N-terminal, R1 and R2 repeats, proline rich region (P) and choline binding domain (CBD) highlight. The approximate boundaries of the fragment PspCN are indicated.

6.1 Binding of full-length FH linker constructs to C3b, C3c and C3d

The interaction between FH and C3b is key to regulation of the alternative pathway of complement. It is thought to involve CCPs 1-4 and CCPs 19-20 binding to adjacent sites on C3b, with intervening FH modules playing an architectural or hinge-like role as discussed in Chapter 1. It was therefore of interest to compare the binding of FH WT with that of the two linker mutants, SL and 3xGLY.

6.1.1 Binding to C3b

An initial experiment using plasma-purified FH (purchased from Complement Technology) was performed as a bench mark (Figure 6.2). A concentration series of 0, 0.0313, 0.0625, 0.125, 0.25, 0.5, 1.0 and 2.0 µM FH was flowed over immobilized C3b (see Methods, section 2.6.2) with dual injections for each concentration. A K_D

value of 0.5 ± 0.06 µM was obtained by fitting steady-state binding levels derived from the background-subtracted traces to a one-to-one steady-state binding model, in line with previously obtained values (0.59 ± 0.04 µM) [63].

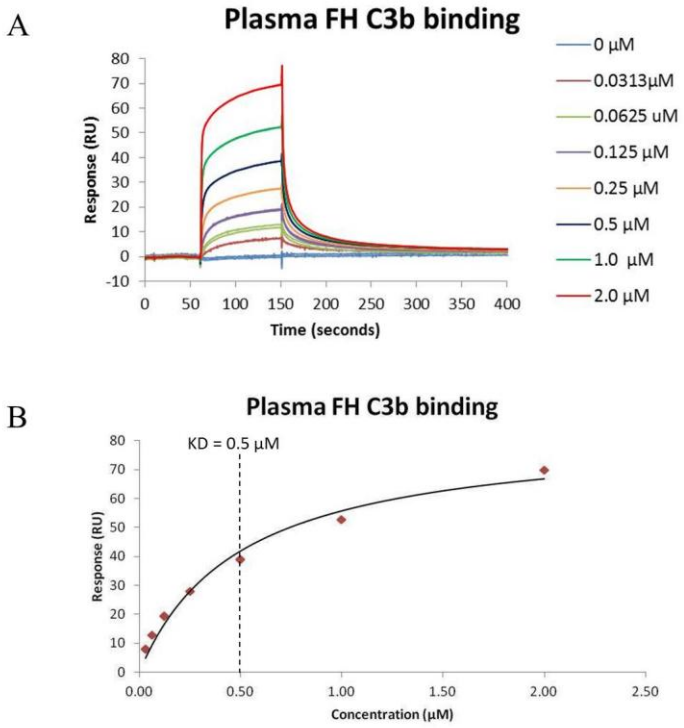


Figure 6.2 Binding of plasma-purified FH to C3b. In (A) is the background-subtracted sensorgram with each of the concentration coloured according to the key. (B) The plot of response obtained *vs.* concentration with the calculated K_D indicated by the dashed line.

Having established that the assays was working by obtaining a K_D for plasma-purified CFH that was in line with expectations, the SPR experiment was repeated with recombinantly produced WT FH (Figure 6.3) instead of plasma-purified material. An identical concentration series was used. A K_D value of 0.58 ± 0.09 µM was obtained, which is very similar to the result for plasma-purified FH in the previous experiment.

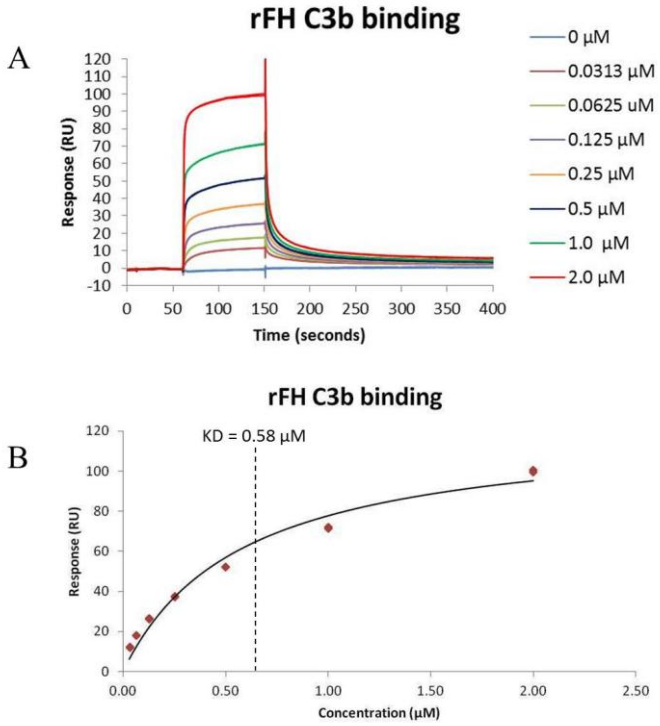


Figure 6.3 Binding of rFH to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the calculated K_D indicated by the dashed line.

The comparable K_D value for FH 3xGLY (Figure 6.4) was estimated at 1.05 ± 0.18 µM. Due to the low concentration of the FH SL stock protein solution, a concentration range of 0, 0.0313, 0.0625, 0.125, 0.25, 0.5 and 1 µM was used and only single injections were performed for each. Despite this, the data fit very well to the one-to-one binding model (Figure 6.5) and the K_D value was reliably estimated at 2.2 ± 0.24 µM.

These affinities of the mutants for C3b are significantly, but not dramatically, different from that of FH WT (~ 0.5 µM) and therefore replicate experiments would be needed to confirm that FH 3xGly and FH SL really do bind two-fold and four-fold, respectively, more weakly to C3b. It is, nonetheless, intriguing that the K_D value estimated for FH SL:C3b is very similar to that (~ 2 µM) of FH 19-20:C3b reported previously (FH 1-4:C3b has a weaker K_D of ~ 10 µM); so it could be that the two C3b-binding sites in this mutant (CCPs 1-4 and CCPs 19-20) are unable to bind simultaneously.

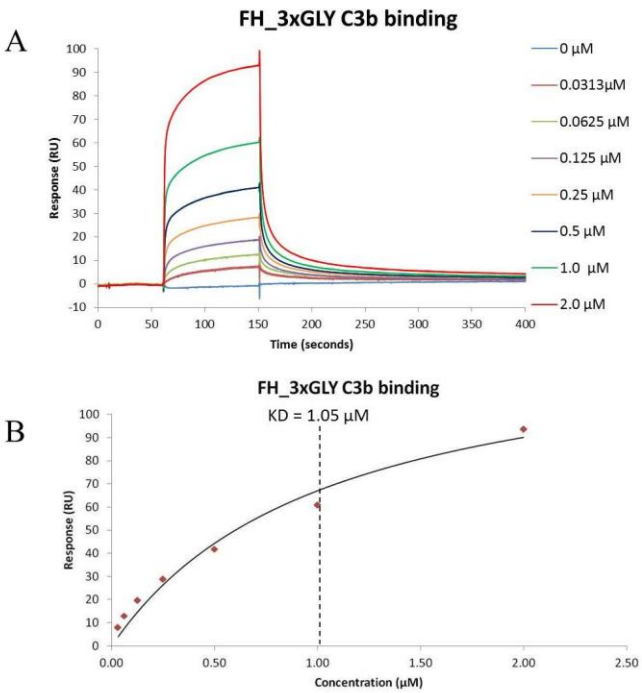


Figure 6.4 Binding of FH 3xGLY to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the calculated K_D indicated by the dashed line.

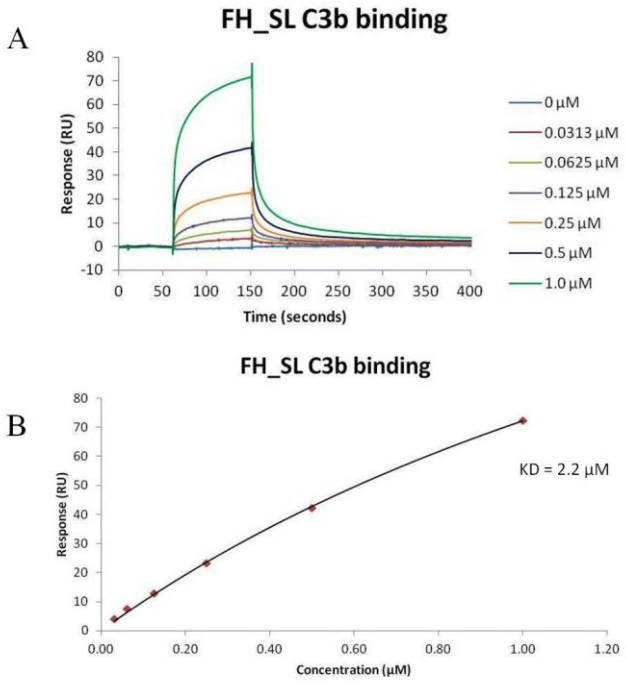


Figure 6.5 Binding of FH SL to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the calculated K_D value.

6.1.2 Binding to C3c

Previous work showed that FH does not bind to C3c, presumably because FI has ablated both the binding site for CCPs 1-4 (due to cleavage within the CUB domain) and for CCPs 19-20 (due to removal of the TED). Using the same concentration series as used for C3b (and a different channel of the same SPR chip – see Methods, section 2.6.2) it was found that none of the versions of FH in the current study bound significantly to C3c (Figure 6.6).

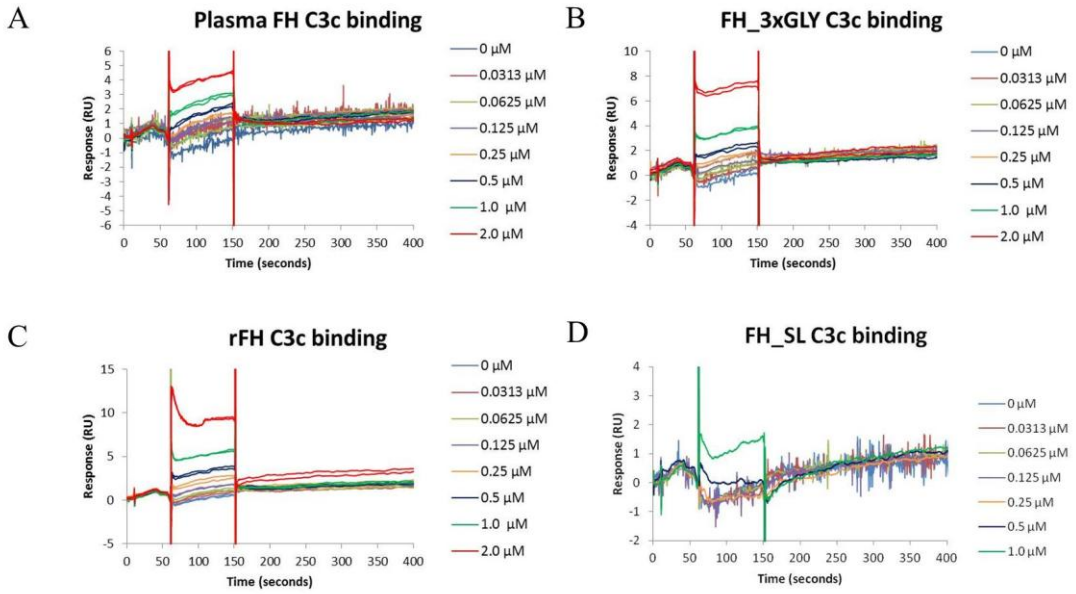


Figure 6.6 Binding of FH linker constructs to C3c. Each sensorgram shows the background-subtracted curves with the concentrations used coloured according to the keys. (A) Plasma-purified FH, (B) FH 3xGLY, (C) recombinant FH (WT) and (D) FH SL.

6.1.3 C3d binding

Previous work from several laboratories suggested that FH binds only weakly to C3d, unlike the FH fragment FH 19-20 that binds to both C3b and C3d with an affinity of ~ 2 µM. An identical concentration series of FH as employed for C3b binding was flowed over a different channel of the same chip that had been loaded with 69 RUs of C3d.

Initial experiments with commercially prepared, C3b-derived C3d (from Complement Technology) and plasma-purified FH showed (Figure 6.7) a very small response. A very rough estimate of the K_D value (13 ± 5 µM) could be extracted

from the data, but the very low number of response units obtained might reflect binding of FH to a small quantity of contaminant on the chip (such as iC3b as might arise from incomplete digestion during the preparation of C3d). Alternatively, the slight leveling off of the curve perceptible in Figure 6.7B might be an artifact and hence the K_D could in fact be much higher than 13 µM.

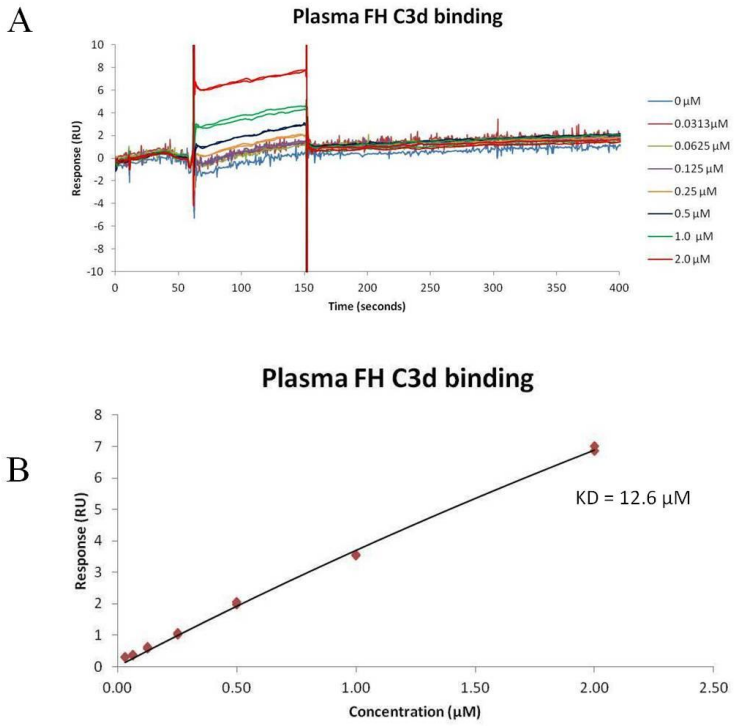


Figure 6.7 Binding of plasma FH to C3d. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the estimated K_D value that cannot be considered reliable in this case due to the low responses obtained and the maximum concentration of FH used.

The three recombinant versions of FH (WT, SL and 3xGLY) all yielded higher responses (than plasma-purified FH) when flowed over immobilized C3d. Yet in none of the response *vs*. concentration curves is there any sign of saturation being approached (Figure 6.8). Thus, no meaningful K_D values could be obtained. The interactions appear, therefore, to be non-specific and one suggestion is that these arise due to the lack of glycosylation (beyond single GlcNAcs) of the recombinant proteins.

In summary, the current results are in full agreement with previous observations that full-length FH, whether recombinant or derived from plasma, binds at least a couple of orders of magnitude less well to C3d than does the C-terminal fragment FH 19-20.

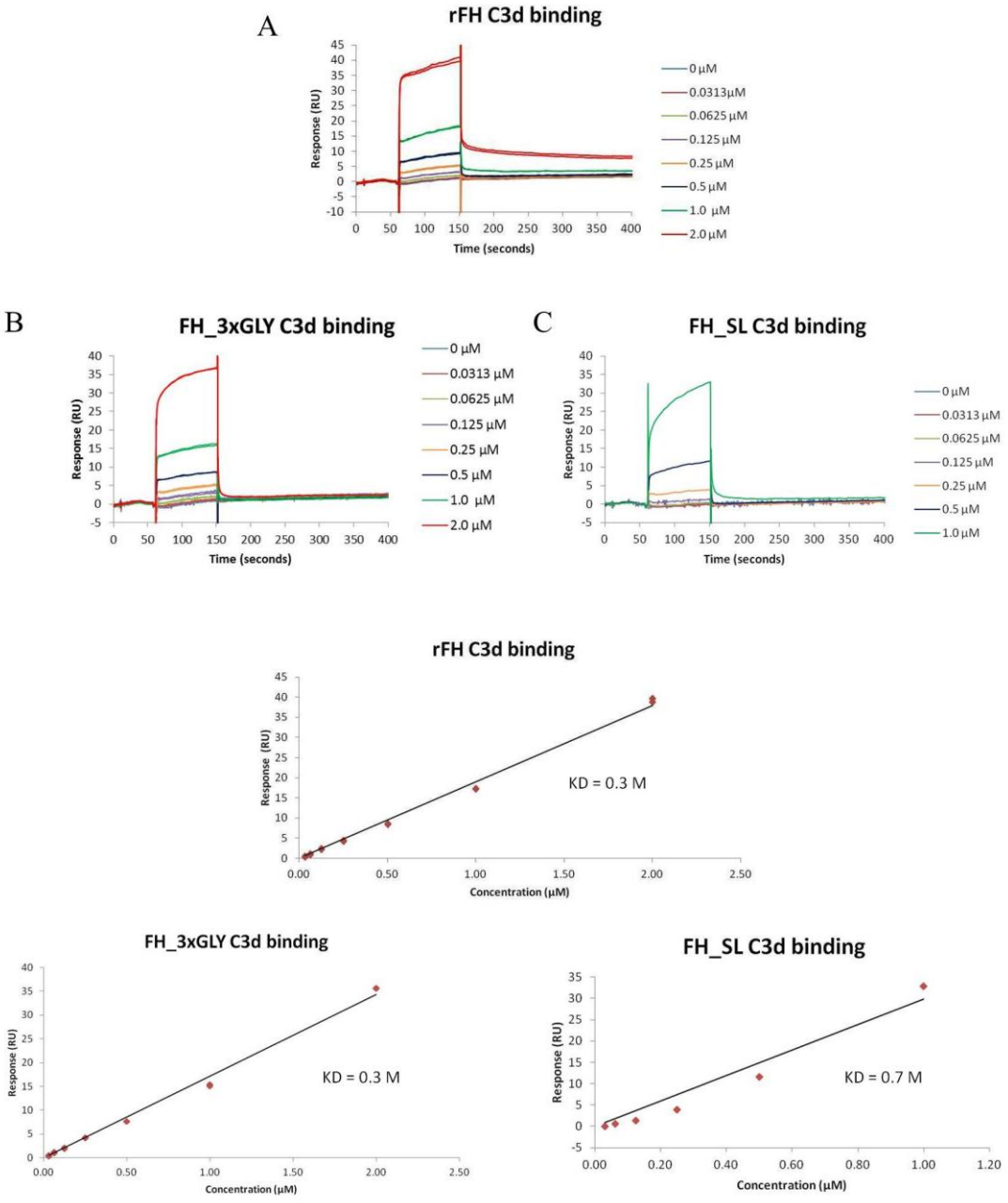


Figure 6.8 FH WT and FH with mutations in the 12-13 linker do not bind significantly to C3d. Each sensorgram shows the background-subtracted curves with the concentrations used coloured according to the keys. The plots of response obtained *vs.* concentration, with estimated K_D values (that are meaningless in all three cases due to the lack of any hint of saturation) are shown below.

6.2 Decay acceleration by full-length FH and linker mutants

The results of the above-described study indicate that all of the tested versions of full-length FH bound to C3b with K_D values in the range of 0.5 to 2 μM. It was important to establish whether mutating the linker had any influence on the C3b.Bb decay-accelerating activity of FH, which is assumed to depend upon initial engagement of FH with C3b (as well as Bb). It was hypothesized that if the SL mutant genuinely binds to C3b twofold less well than FH 3xGLY, and fourfold less well than FH WT, a matching deficiency in DAA might occur.

At 2.5 nM FH (or 2.5 nM FH mutant) the results showed (Figure 6.9; see Methods, section 2.6.3, for a description of how this experiment was preformed and the data processed) that SL was indeed the poorest of the recombinant proteins tested while the recombinant FH WT and FH 3xGLY appeared to have virtually indistinguishable activities. The plasma-purified FH performed less well in this assay than recombinant FH as was observed previously [140].

When the concentrations of the FH constructs were increased to 5 nM and then 10 nM this trend persisted, with FH SL displaying the lowest activity of the four proteins tested. As at the lower concentration, there is only a small difference in the rate of activity between WT and 3xGLY and both have a higher activity than plasma FH.

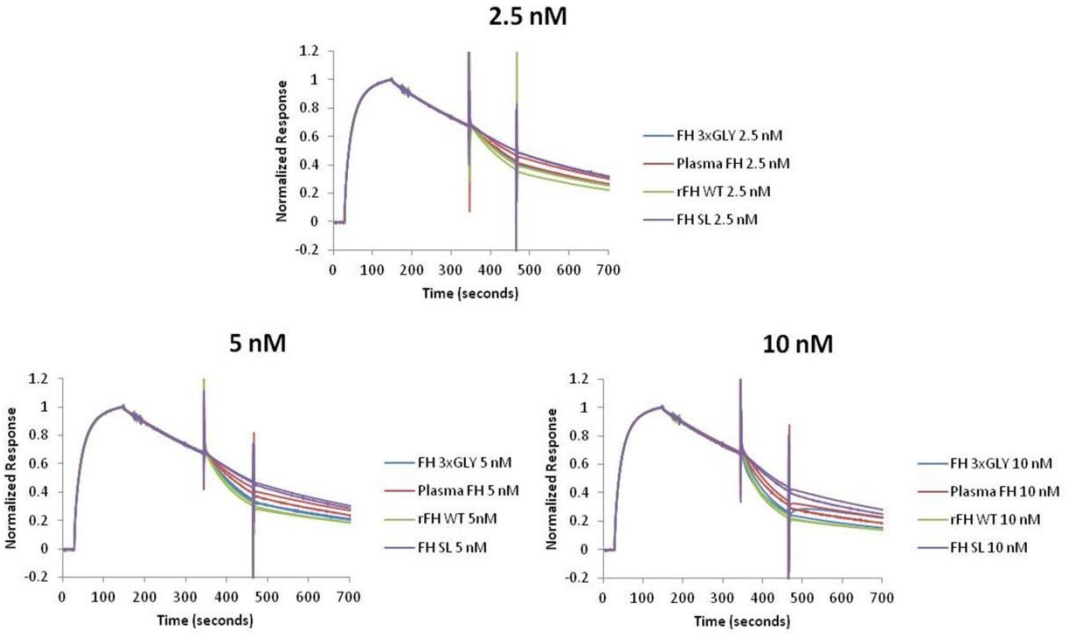


Figure 6.9 Decay-accelerating activities of various versions of full-length FH. The sensorgrams were obtained according to the procedures outline in the Methods (section 2.6.3) and show initial assembly of the C3 convertase (C3b.Bb) followed by its natural decay and then (first spike) addition of FH that gives rise to a faster dissociation rate. Finally (second spike) injection of FH ceases and the inherent rate of decay resumes.

6.3 Interaction of FH linker constructs with PspC of S. pneumoniae

In order not to be destroyed by the complement system, many pathogens have developed means of avoiding deposition of C3b on their surfaces [110] (as outlined in the Introduction). For example, one of the ways *S. pneumoniae* avoids complement-mediated destruction is by expressing the gene for the surface-exposed protein PspC, which is a virulence factor that can bind to FH [112]. Previous work on the N-terminal segment of PspC (called PspCN) has shown that this is the region required to bind FH [183], and that within FH the binding site of PspCN is located within the central CCPs 8-10 [184], *i.e.* well away from the C3b-binding sites in CCPs 1-4 and 19-20. It is therefore possible that PspC binding alters the architecture of FH and modulates its C3b-binding properties and complement-regulatory functions.

As reported below, the various versions of FH 8-15 and full-length FH described in the previous Chapter were used to study how mutations in the CCPs 12-13 linker affect the binding of PspCN. In further studies, the effects of PspCN on the functions of the FH linker mutants were investigated.

6.3.1 Binding of linker constructs to PspCN

Previous work had shown that native-sequence FH 8-15 (WT) binds tightly to PspCN with a K_D of 1.7 nM [184], and so initial experiments aimed to replicate this result. A concentration range of 1, 2, 4, 8, 16, 32 and 64 nM FH 8-15 WT was flowed over His-tagged PspCN attached to the chip (see Methods, section 2.6.1) with a double injection being performed at 64 nM.

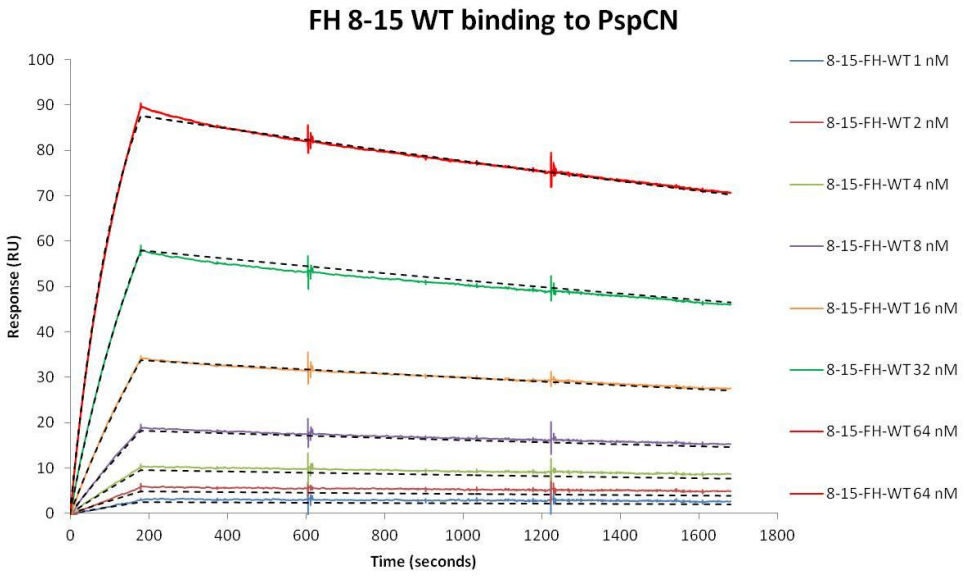


Figure 6.10 Binding of FH 8-15 WT to His-tag PspCN. Background-subtracted traces are shown for all concentrations and are coloured as indicated in the key. The dashed lines represent the fitted data for each concentration.

The sensorgrams for FH 8-15 WT (Figure 6.10) showed a rapid association rate (k_{on}) and a very slow dissociation rate (k_{off}), as was observed in previous work. When the kinetics data from the subtracted traces was fitted to a 1:1 Langmuir model, a K_D value of 1.3 nM was obtained that is in very good agreement with the previously obtained value [184]. A further study was performed to ascertain how the linker modifications affected the affinity of FH 8-15 for PspCN.

Using identical concentrations of FH 8-15 3xGLY the sensorgrams (Figure 6.11) again showed fast association and slow dissociation. The K_D value obtained (from fitting the subtracted traces to a 1:1 Langmuir model) was 0.7 nM, indicating that binding of this mutant to PspCN was at least as good as that of FH 8-15 WT.

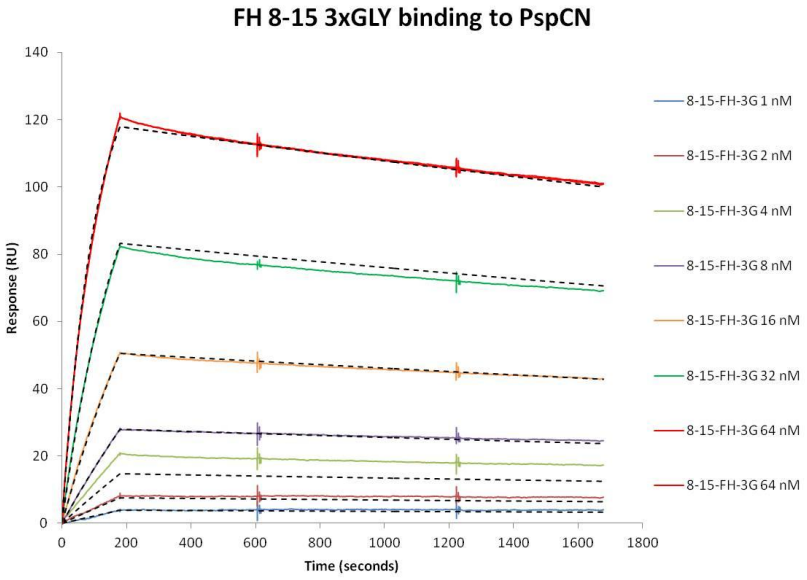


Figure 6.11 Binding of FH 8-15 3xGLY to His-tag PspCN. Background-subtracted traces are shown for each concentration and are coloured as indicated in the key. The dashed lines represent the fitted data for each concentration.

In a further SPR experiment, an identical concentration series of FH 8-15 SL was flowed over PspCN. Again, rapid association and slow dissociation were observed (Figure 6.12), with an estimated K_D of 0.5 nM. Although the data does not fit quite as well to the binding model as for the other constructs, it is reasonable to conclude that this mutant binds at least as well as FH 8-15 WT and FH 8-15 3xGLY to PspCN.

In summary, these results are consistent with the binding of PspCN to CCPs 8-10 within FH 8-15 independently of the CCPs 12-13 linker. The results are inconsistent with a model in which CCPs 8-15 wrap around PspCN to form the stable complex. The observation that FH 8-15 binds about 30-fold better than FH 8-9 [184] therefore suggests that CCP 10 additionally interacts with PspCN (and this was supported by cross-linking studies performed by Zhou Chen, University of Edinburgh, personal

communication.) while participation of CCPs 11 or 12 is also possible; a FH 8-10 construct is needed to test these possibilities.

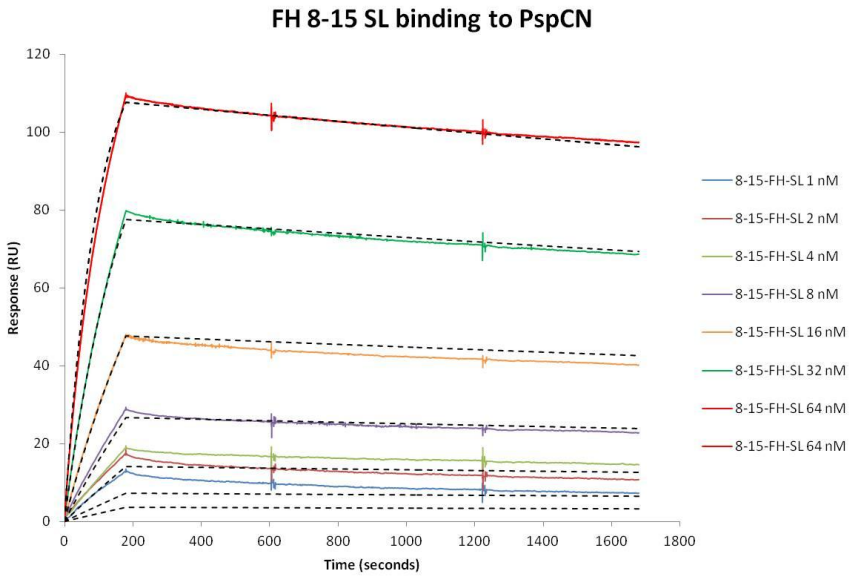


Figure 6.12 Binding of FH 8-15 SL to His-tag PspCN. Background-subtracted traces are shown for each concentration and are coloured according to the legend. The dashed lines represent the fitted data for each concentration.

To examine the effects of the linker mutations in full-length FH, initial SPR experiments were performed with plasma-purified FH (purchased from Complement Technologies). The concentration series used in this case was 0.25, 0.5, 1, 2, 4 and 8 nM FH with 8 nM FH being injected twice.

As expected (Figure 6.13) rapid association is followed by barely detectable dissociation. The K_D was estimated, by fitting the background-subtracted data to a 1:1 Langmuir model, at 0.05 pM which is outside the range that can be accurately measured but consistent with an extremely tight interaction (*i.e.* effectively irreversible binding).

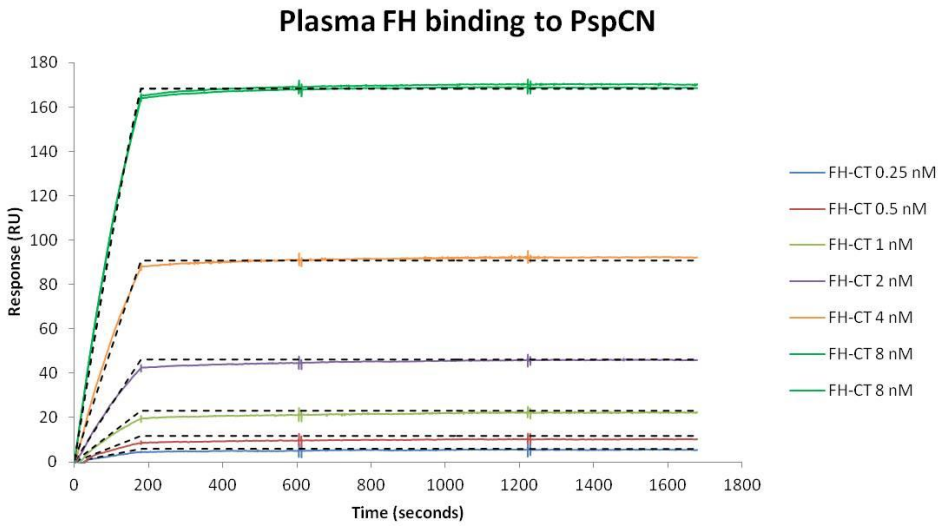


Figure 6.13 Binding of plasma-purified FH to His-tag PspCN. Background-subtracted traces are shown for each concentration and are coloured according to the key. The dashed lines represent the fitted data for each concentration.

Virtually identical results were obtained with recombinant FH (Figure 6.14), although the fit of the data to the model was poor, particularly for the highest concentration of FH. This is the appropriate control protein for the mutants of FH produced in the current study.

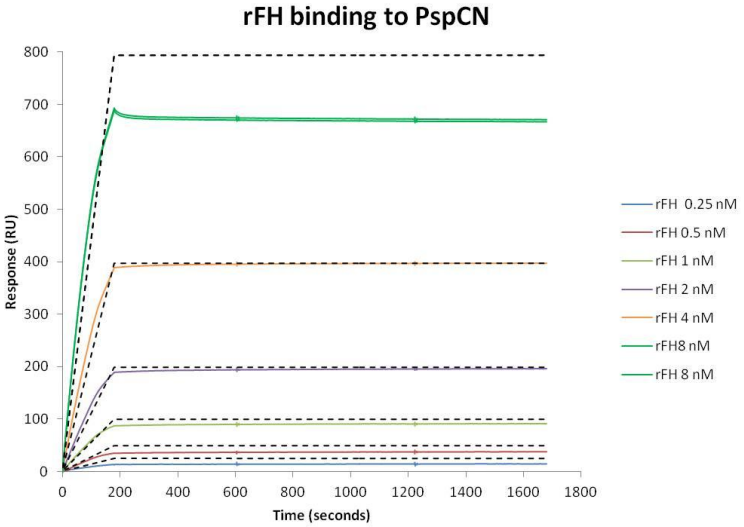


Figure 6.14 Binding of recombinant FH WT to His-tag PspCN. Background-subtracted traces are shown for each concentration and are coloured according to the key. The dashed lines represent the fitted data for each concentration. Poor fits were obtained for some of the traces but this did not affect the overall conclusion that recombinant FH binds to PspCN with approximately the same affinity as plasma-purified FH does.

The sensorgrams (Figure 6.15) for the equivalent experiment carried out with FH 3xGLY again showed effectively irreversible binding to PspCN. The K_D value obtained was 10-fold weaker at 0.45 pM, but it is hard to ascribe much meaning to such a value since these affinities are outside the range that can be accurately determined by SPR.

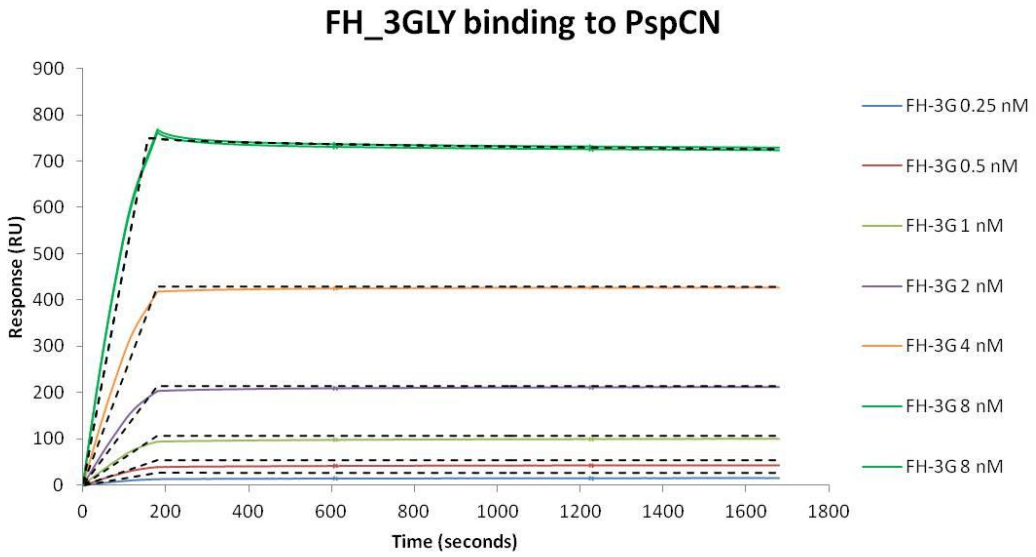


Figure 6.15 Binding of FH 3xGLY to His-tag PspCN. Background subtracted traces are shown for each concentration and are coloured as indicated in the key. The dashed lines represent the fitted data for each concentration.

The results for FH SL (Figure 6.16) suggested an increase in the dissociation rate compared to that seen for FH WT and FH 3xGLY. The complex formed with PspCN by this linker mutant is not irreversible; dissociation is perceptible over the 30-minute course of data collection. In this case a K_D of 80 nM was estimated although the data does not fit particularly well to the binding model so there is a large error associated with this value.

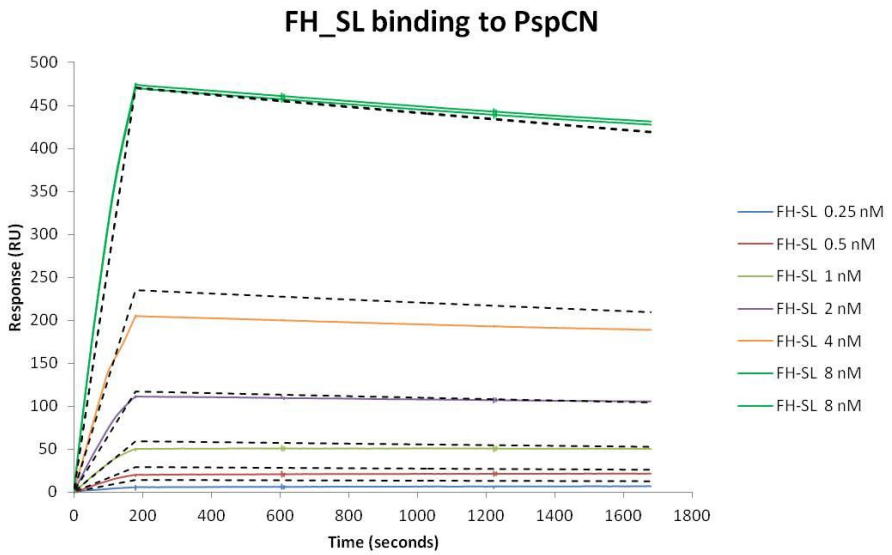


Figure 6.16 Binding of FH SL to His-tag PspCN. Background-subtracted traces are shown for each concentration and are coloured as indicated in the key. The dashed lines represent the fitted data for each concentration.

In summary it can be concluded that making the FH 12-13 linker more flexible by substituting bulky residues with glycine residues does not prevent FH from binding very tightly to PspCN, and much more tightly than FH 8-15. This implies that modules other than CCPs 8-15 stabilize the FH:C3b complex. Since PspCN is quite small (~110 residues) it seems unlikely that more than three or four CCPs (60 residues each) make simultaneous direct contact with the bacterial polypeptide. More likely is that the PspCN stabilizes a conformation of FH in which non-sequential CCPs participate in complex-stabilizing interactions. This confirmation is unavailable to FH SL that binds no better than FH 8-15 SL.

Table 6.1 Summary of K_D values obtained for FH binding to PspCN

FH CONSTRUCT	K_D VALUES (M)
8-15 WT	1.25×10^{-9}
8-15 3xGLY	0.72×10^{-9}
8-15 SL	0.45×10^{-9}
Full length WT	0.05×10^{-12}
Full length 3xGLY	0.45×10^{-12}
Full length SL	80×10^{-9}

6.3.2 Binding of PspCN:FH complexes to C3b, C3c and C3d.

Previous work had suggested that the FH:PspCN complex has a higher affinity for C3b and C3d than FH alone [184]. Using this observation as a starting point it was decided to investigate how PspCN could alter the binding to the C3 fragments of the full-length FH linker mutants.

C3b binding

A complex of FH and PspCN (FH:PspCN) was formed, using a 2:1 molar mixture of the bacterial protein and plasma-purified FH, and this was flowed over the same C3b-loaded sensorchip surface (Figure 6.17) that was used for studies of FH in the absence of PspCN (section 6.1.1, Figure 6.2). The K_D value obtained was 0.16 ± 0.02 µM, which is approximately three times stronger than that observed for FH alone (*i.e.* no PspCN) binding to C3b. This is entirely consistent with previous (unpublished) observations made by other members of the Barlow lab.

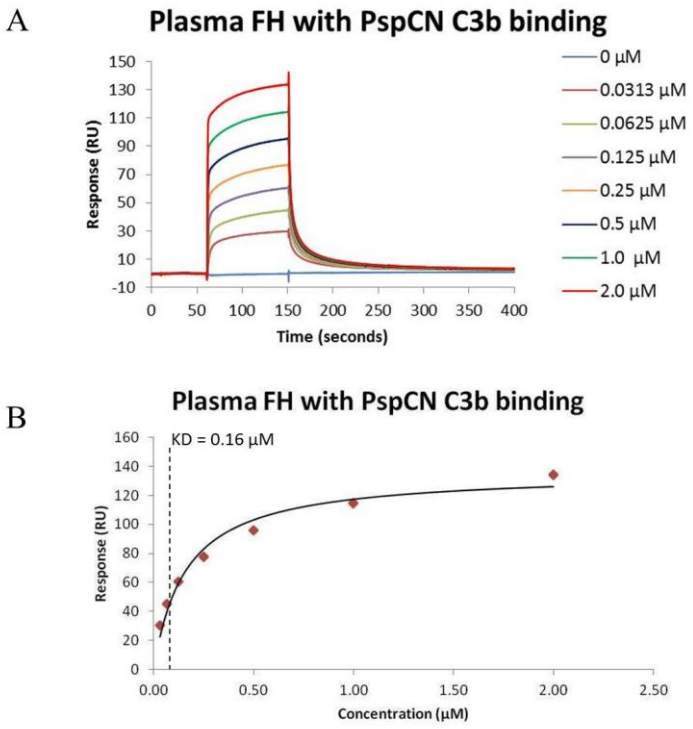


Figure 6.17 Binding of plasma-purifed FH:PspCN complex to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the calculated K_D indicated by the dashed line.

The K_D value for recombinant FH:PspCN complex binding to C3b (Figure 6.18) was determined in the same way, and found to be 0.16 ± 0.02 µM - identical to the plasma-purified:PspCN complex and 3.5 times stronger than recombinant FH alone (Figure 6.3).

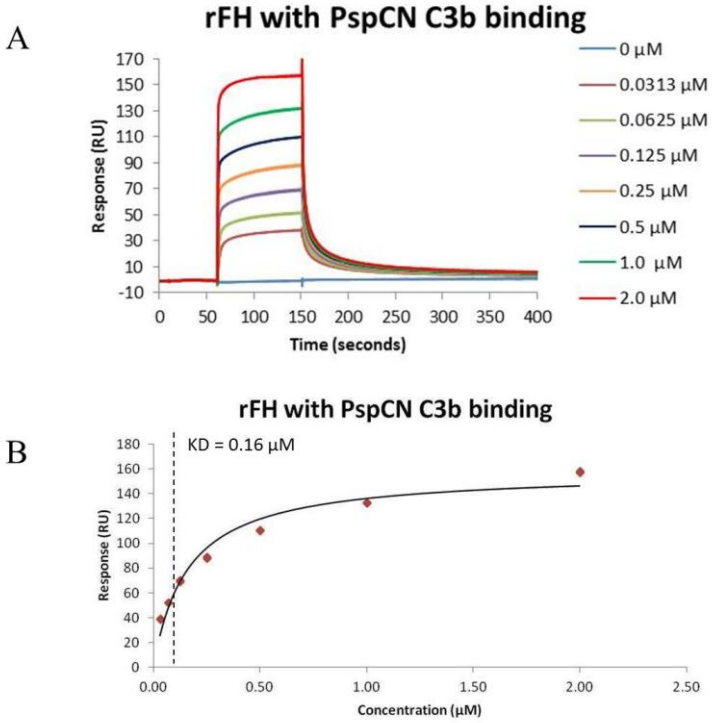


Figure 6.18 Binding of recombinant FH:PspCN complex to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the calculated K_D indicated by the dashed line.

FH 3xGLY:PspCN and FH SL:PspCN complexes were prepared in the same way as the wild-type FH:PspCN complexes and flowed over the same C3b-loaded sensorchip surface. Note that in the case of FH SL, the injections at the lower concentrations (~31, 62 and 125 nM) may in fact contain a mixture of free FH and PspcN-complexed FH due to the higher K_D for the FH SL:PspN complex (~ 90 nM). FH 3xGLY:PspCN (Figure 6.19) and the FH SL:PspCN bound to C3b (Figure 6.20) with K_Ds of 0.24 ± 0.04 µM and 0.42 ± 0.06 µM, respectively. These values correspond to a very roughly fourfold increase in affinity for C3b compared to FH 3xGLY or FH SL on their own. Note that the data for FH SL:PspCN fit well to the model despite the much weaker complex formed by these proteins.

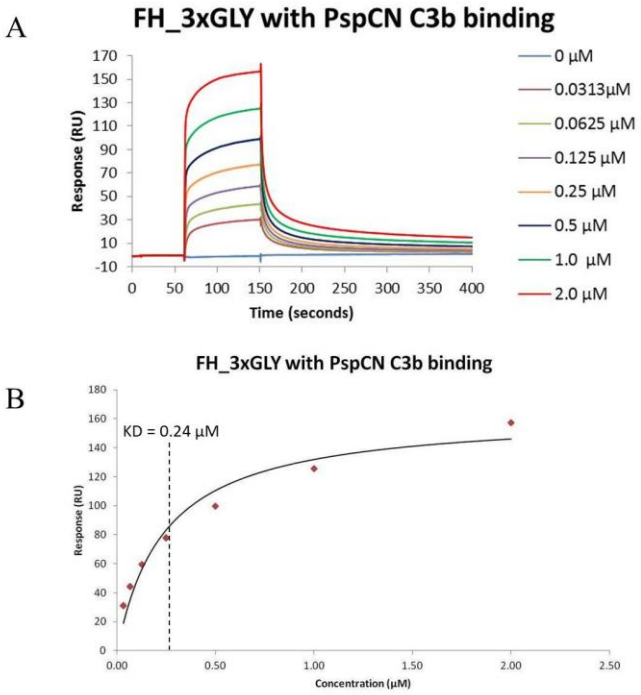


Figure 6.19 Binding of FH 3xGLY:PspCN complex to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs*. concentration with the calculated K_D indicated by the dashed line.

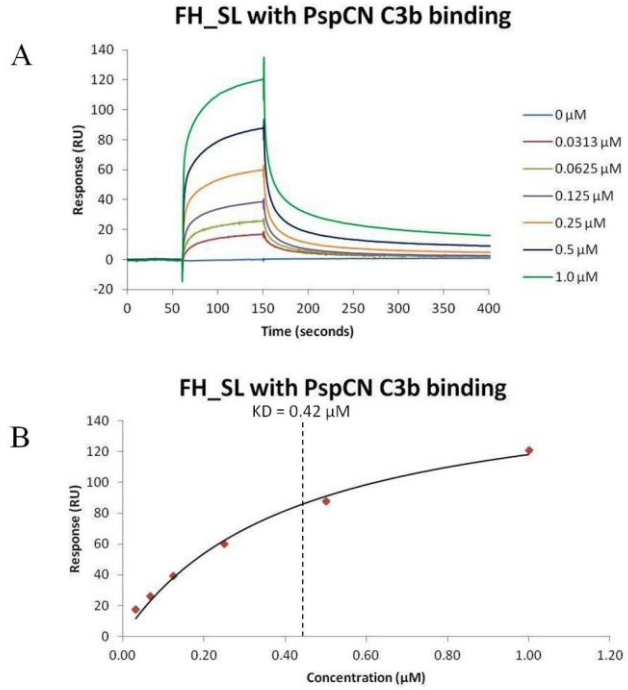


Figure 6.20 Binding of FH SL:PspCN complex to C3b. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of

response obtained *vs.* concentration with the K_D indicated by the dashed line.

In summary (Table 6.2) these results deliver a remarkably consistent story in which the binding of PspCN brings about an approximately fourfold improvement in affinity for C3b. The lower affinity of FH SL for PspCN did not seem to have an impact here, perhaps because the FH SL:PspCN K_D was overestimated in the previously described SPR-based calculation (Figure 6.5); visually those data look similar to the date for FH 8-15 SL:PspCN (K_D = ~1 nM, Table 6.1)

Table 6.2 Summary of K_D values obtained for all FH construct binding to C3b with and without the presence of PspCN.

FH construct	FH only (μM)	FH:PspCN complex (μM)
Plasma	0.5 ± 0.06	0.16 ± 0.02
WT	0.58 ± 0.09	0.16 ± 0.02
3xGLY	1.05 ± 0.18	0.24 ± 0.04
SL	2.2 ± 0.24	0.42 ± 0.06

C3c binding

The C3c-binding experiments performed on the FH variants were repeated for the equivalent FH:PspCN complexes (Figure 6.21). There was some evidence of binding to the immobilized C3c on the chip, but it was characterized by fast off and on rates and low responses, suggesting weak or non-specific binding. There were significant differences between the proteins tested, with FH SL:PspCN showing the smallest response, while recombinant FH:PspCN produced the biggest response. It is possible that this surface was contaminated with a small quantity of uncleaved C3b or iC3b.

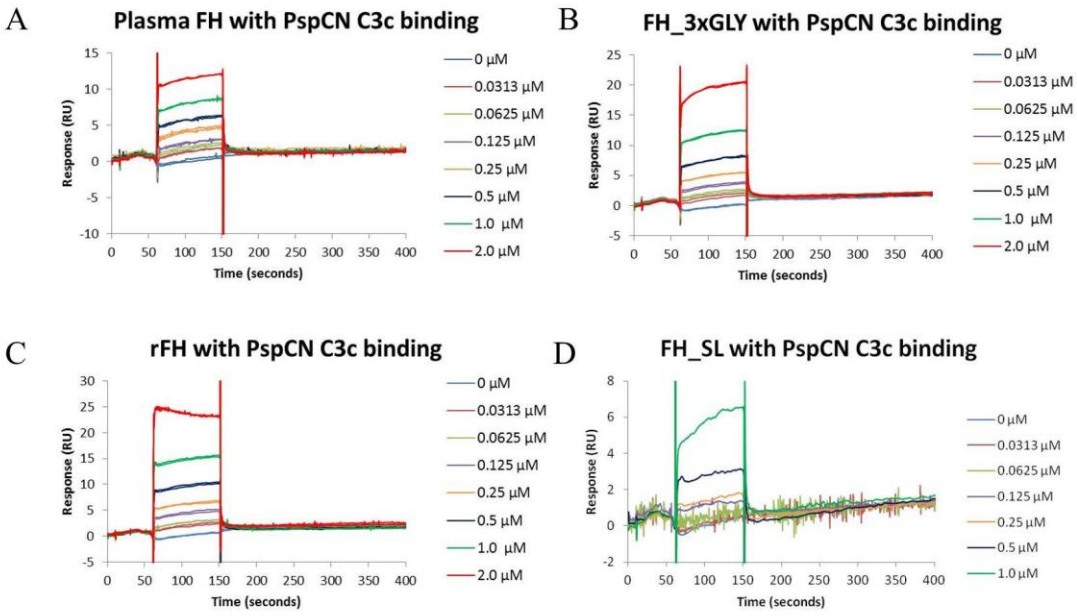


Figure 6.21 Binding of FH:PspCN complexes to C3c. Each sensorgram shows the background-subtracted curves with the concentrations used coloured according to the keys. (A) plasma-purified FH:PspCN, (B) FH 3xGLY:PspCN, (C) FH WT:PspCN and (D) FH SL:PspCN.

C3d binding

Upon the addition of PspCN to form the FH:PspCN complex (Figure 6.22) a K_D value of 2.3 ± 0.2 µM was obtained for C3d. This exceeds the highest concentration used for the SPR experiment (2 µM) but the data fit well to the model and this calculated value is probably quite reliable. This is dramatically better than the binding of FH alone (for which only a tiny response was recorded and no K_D could be accurately determined). This is in line with previously obtained results, and is consistent with the aforementioned enhanced affinity of FH:PspCN for C3b. Importantly, it suggests that PspCN reveals a previously masked or cryptic C3b/d-binding site of FH, presumably in CCPs 19-20.

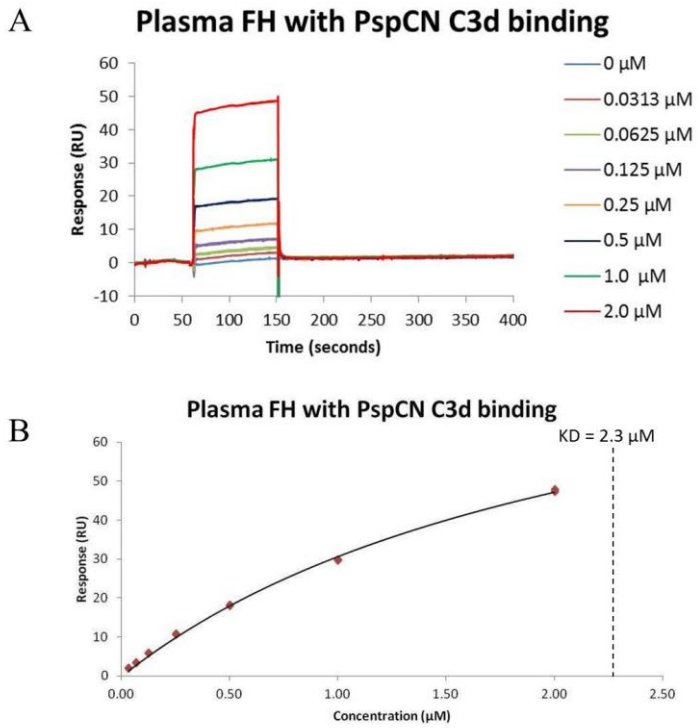


Figure 6.22 Binding of plasma FH:PspCN complex to C3d. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained vs. concentration, with the calculated K_D. Despite this being greater than the maximum concentration used in the series, the data fit well to the curve, and the calculation of K_D. is probably reliable.

The recombinant FH:PspCN complexes all showed an improvement in binding to C3d. For FH WT (Figure 6.23) the K_D was 2.1 ± 0.2 µM which is very similar to the result seen for the plasma-purified FH:PspCN complex, although the response obtained in this case was significantly higher.

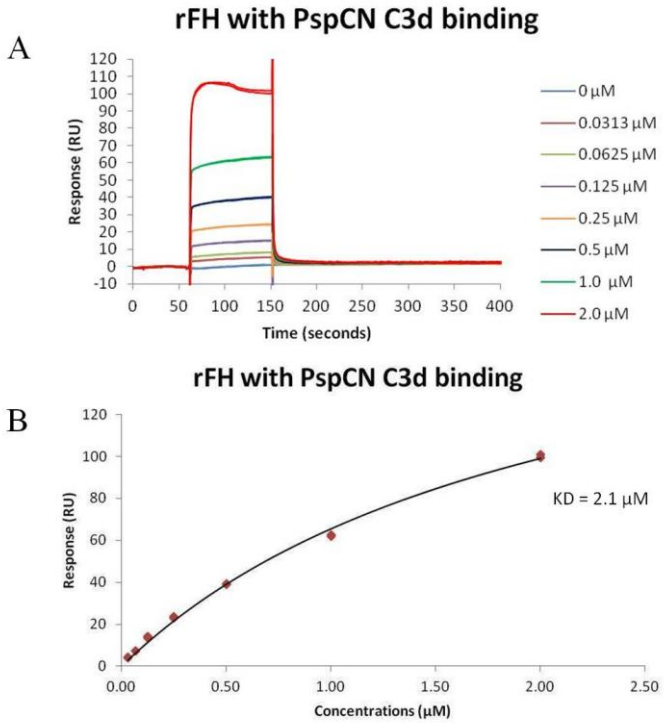


Figure 6.23 Binding of WT FH:PspCN complex to C3d. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs*. concentration along with the calculated K_D.

For the FH 3xGLY:PspCN complex binding to C3d, the response was high, but the calculated K_D suggested weaker binding (compared to WT) at 4.5 ± 0.7 µM (Figure 6.24) that was well in excess of the highest FH concentration used in the series and therefore cannot be regarded as wholly reliable. It is however about half as strong as the value for WT, which is in line with the relative values for FH WT and FH 3xGLY binding to C3b (0.5 µM and 1 µM).

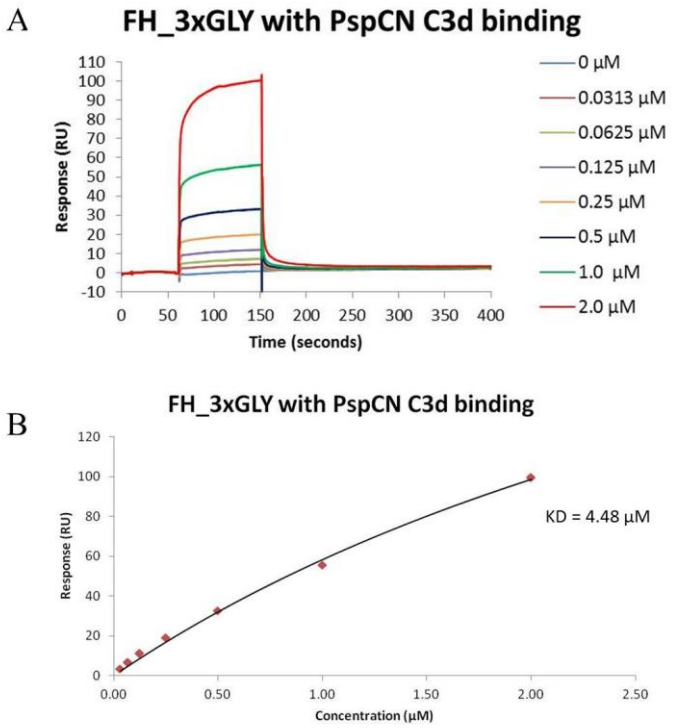


Figure 6.24 Binding of FH 3xGLY:PspCN complex to C3d. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the legend. (B) shows the plot of response obtained *vs*. concentration with the calculated K_D.

Interestingly, in the case of the FH SL:PspCN complex (Figure 6.25) (unlike FH WT and 3xGLY) binding to C3d was much weaker with no sign of approaching saturation at 2 µM and therefore a K_D value in excess of several 10s of µM.

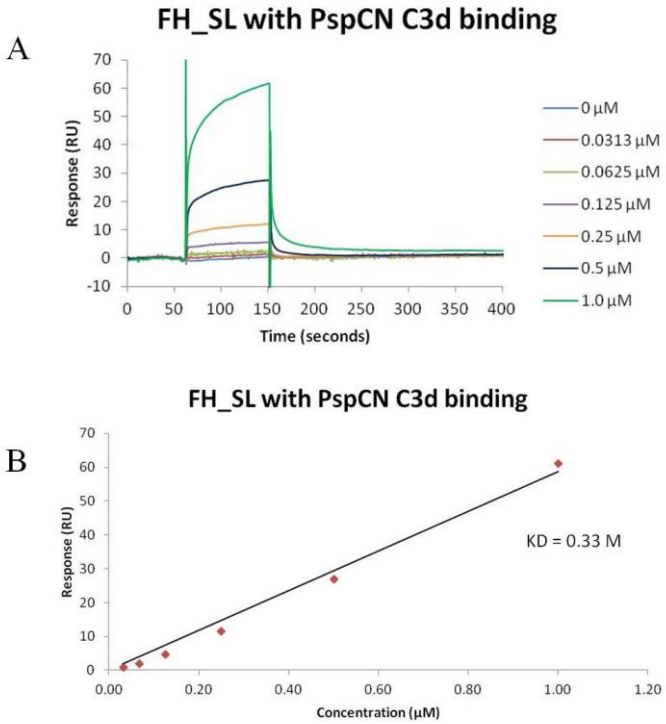


Figure 6.25 Binding of FH SL:PspCN complex to C3d. In (A) is the background-subtracted sensorgram with each of the concentrations coloured according to the key. (B) shows the plot of response obtained *vs.* concentration with the calculated K_D. This number is meaningless due to the lack of any sign of saturation at 2 µM.

In summary (Table 6.3), FH on its own does not bind with a physiologically meaningful K_D to C3d. In contrast when PspCN is present both the WT and 3xGLY samples bind to C3d about as well as FH 19-20.

Table 6.3 Summary of K_D values obtained for all FH construct binding to C3d with and without the presence of PspCN.

FH construct	FH only	FH:PspCN complex
Plasma	12.6 ± 4.27 µM	2.32 ± 0.16 µM
WT	-[A]	2.10 ± 0.20 µM
3xGLY	-[A]	4.48 ± 0.69 µM
SL	-[A]	-[A]

–[A] means that the K_D was too high to be estimated based on current data.

6.3.3 Decay-accelerating activity of FH:PspCN complexes.

From studies of FH and its linker mutants binding to C3b and its fragments, it is clear that PspCN enhances the affinity between these proteins. The next step was to investigate whether the FH:PspCN complexes also increased the rate of decay of the C3 convertase more than FH in isolation. As before, the complexes were formed using a 2:1 ratio (PspCN:FH). At the lowest concentration of 2.5 nM (Figure 6.26) it is clear that the rate of decay for all of the FH variants has increased compared to just FH SL alone. The enhancements seen at 2.5 nM were also evident in assays with 5 nM and 10 nM FH:PspCN complex. The presence of PspCN clearly boosts the ability of all the FH constructs to decay the C3 convertase.

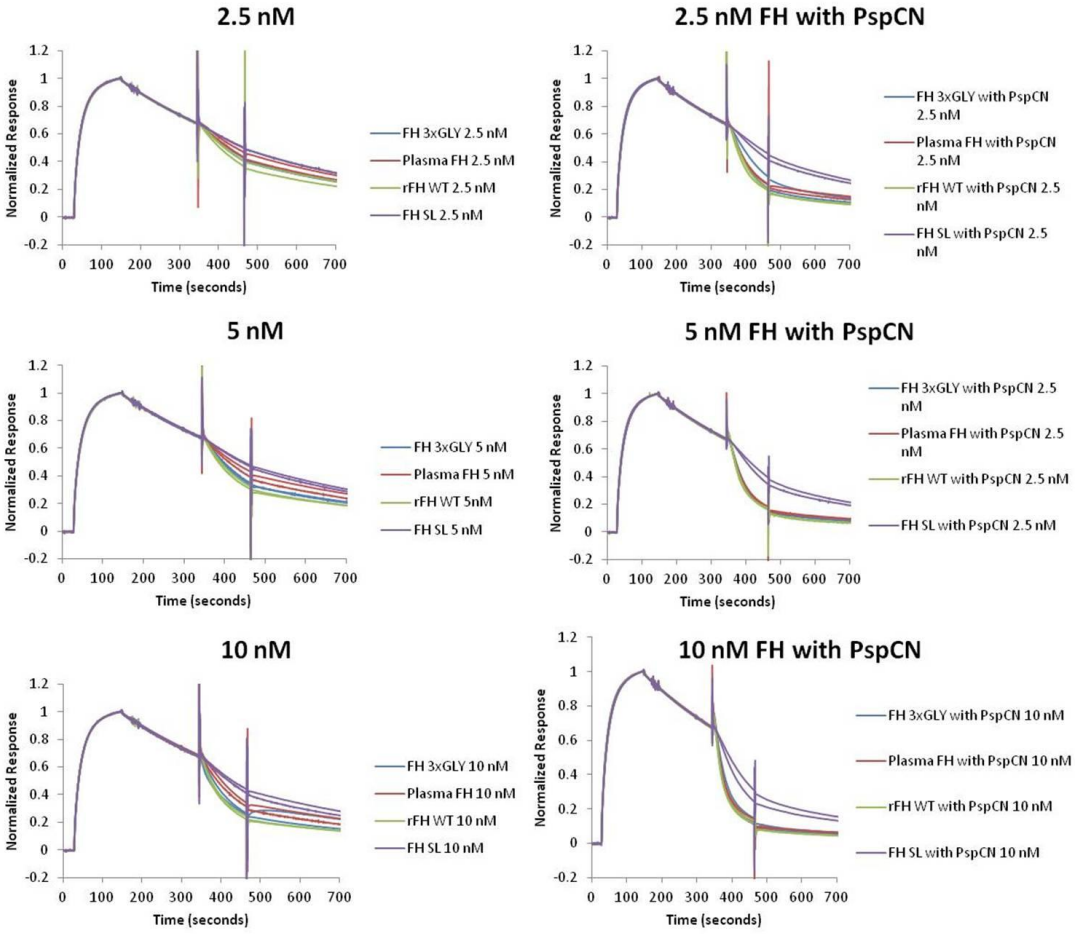


Figure 6.26 Decay-accelerating activity of FH:PspCN complexes. The sensogram have each had the background subtracted and proteins are coloured according to the keys. The FH-only sensograms (2.5 nM etc) are shown alongside (and to the left of) those for the FH:PspCN complexes (2.5 nM *etc.* FH with PspCN).

6.4 Discussion

Both plasma-purified and recombinant, native-sequence, full-length FH bound to amine-immobilized C3b with SPR-derived K_D values in the range of 0.5-0.6 μM. Both versions of FH had very low affinities for C3c and C3d. Thus the C3d-binding site of FH in CCPs 19-20 must be hidden or "cryptic".

Both CCPs 12-13 linker mutants (FH 3xGLY and FH SL) bound significantly less well to C3b although the differences in affinities (twofold and fourfold, respectively) were not radical. This is not surprising given that the C3b-binding regions at either terminus of FH remain intact and are at least five modules away from the engineered linker. Like wild-type FH, neither mutant bound significantly to C3d or C3c. It was notable that the affinity of FH SL for C3b equates with that of CFH 19-20 alone. This may indicate that a version of FH with a short, relatively stiff linker between CCPs 12 and 13 is unable to bind to its two adjacent recognition sites on C3b simultaneously and thus there is no avidity effect. Such a notion is consistent with a "sprung open" conformation of FH SL, and it is in agreement with the SAXS and NMR data for FH 8-15 SL presented in the previous chapter. On the other hand, the low affinity of FH SL for C3d indicates that the C-terminal C3d binding site remains cryptic as it is in the wild-type protein

All variants of FH 8-15 bound tightly (~ 1-2 nM) to PspCN, consistent with a model (based on additional evidence, not shown) in which PspCN (~12 kDa) contacts CCPs 8-10 only. Full-length native-sequence FH (recombinant or plasma-purified) bound several orders of magnitude more strongly than FH 8-15 to produce complexes that are irreversible on a biological time scale. The actual K_D values for these complexes of native-sequence FH with the bacterial protein are hard to assess due to their extremely slow decay rates. A working hypothesis is that the additional CCPs in full-length FH (CCPs 1-7 and 16-20) participate in intermodular contacts that favor complex formation and stability. A similar observation was made for FH 3xGLY but on the other hand it was striking that FH SL did not form an irreversible complex with PspCN and its affinity for the bacterial protein was no better than that of FH 8-15 SL (the calculated K_D for FH SL:PspCN was in fact 40-fold weaker than that for

FH 8-15 SL:PspCN although visually, the sensorgrams were very similar). Thus, in this mutant, the contracted CCPs 12-13 linker prevents the formation of a stable complex with PspCN, presumably by disfavoring a conformation that allows the intermodular interactions alluded to earlier.

All of the variants, when in complex with PspCN, bound approximately threefold to fourfold better to C3b. Strikingly, PspCN enhances the affinity of plasma-purified and recombinant FH for C3d by two orders of magnitude. A similar result was obtained for FH 3xGLY but – interestingly - not for FH SL that displayed little affinity for C3d. In line with the previously suggested hypothesis (above) these observations suggest that in the PspCN-stabilized complex. FH has adopted a conformation that binds well to C3b. An obvious possibility is that in the complex (with FH WT or FH 3xGLY), N-terminal and C-terminal C3b-binding sites are fully revealed and juxtaposed appropriately for simultaneous engagement with the two adjacent recognition regions on C3b. The revelation of the TED/C3d-binding site is not possible in FH SL because the engineered linker prevents the required orientation of the modules (CCPs 12 and 13) on either side and this has repercussions for the overall dynamic architecture of the FH molecule. Nonetheless (see above) FH SL:PspCN is still a better C3b binder than FH SL alone, so some aspects of the PspCN-induced "active" conformation must still be achievable despite the short linker.

The rate of decay of the C3 convertase in the presence of FH and FH mutants, with and without PspCN, was tested in an SPR-based approach that allowed the formation and decay of the C3b.Bb complex to be monitored in real time. The results (in absence of PspCN) showed that FH SL had the lowest activity while recombinant FH WT and FH 3xGLY had similar rates of decay. In these experiments, plasma-purified FH underperformed recombinant FH and was almost as poor as FH SL. Upon addition of PspCN, decay-accelerating activities were enhanced in all four versions of FH but to different extents such that plasma-purified FH and recombinant FH as well as FH 3xGLY all had similar activities, while SL now stands out for its much poorer performance. These results are again consistent with the hypothesis of PspCN stabilizing an active conformation of FH that can bind

optimally to C3b within the C3b.Bb complex to displace Bb (that has an overlapping binding site on C3b). The short linker between CCPs 12 and 13 presumably disfavours adoption of this conformation by FH SL.

Previous work using SPR has narrowed the binding site for PspCN to CCPs 8 and 9 of FH and showed that FH 8-15 has a very tight binding affinity for PspCN with a very slow off-rate. The results reported here showed that complexes of PspCN with the mutants also dissociated very slowly although intriguingly the complex with FH 8-15 SL did appear to decay more rapidly than WT or 3xGLY and the estimated KD for FH 8-15 SL binding to PspCN was accordingly weaker. The poorer fits of the data to the binding model in the case of SL provides grounds for caution and suggest that these experiments need to be repeated.

Next, the engineered linkers were characterized in the context of full-length FH. Again, FH WT and FH 3xGLY both bound effectively irreversibly to PspCN with KD values that were too tight to measure reliably (due to virtually in detectable dissociation rates). In the case of the FH SL:PspCN complex, some dissociation was evident over the course of the sensogram and a KD in the region of 80 nM was calculated. This result is anomalous given that the FH 8-15 SL construct yielded a 2 nM KD. But it clearly highlights the need for preparation of a fresh sample of FH SL and re-recording of the data. In the meantime it remains possible that the SL, despite not disturbing the overall compact nature of the central segment of FH, prevents FH from binding properly to PspCN.

The overall conclusion that can be drawn from these experiments is that the shortened linker prevents CCPs 12 and 13 from adopting their V-shape. This decreases the inherent ability of FH to adopt an active conformation that is stabilised by PspCN (to enhance bacterial protection), and which might also be stabilised by markers on self-surfaces to afford them the protection they need from C3b deposition. These observations are potentially important for understanding the mechanism of FH and designing FH-derived therapeutic inhibitors.

More work, however, is required to firm up the potentially very important reported results: most importantly new, bigger, batches the of the full-length FH SL mutant

and of FH 8-15 SL should be produced and proteolytic clipping during purification (see last chapter) must be avoided. The SAXS data for FH 8-15 WT versus FH 8-15 SL should be re-collected (using glycerol to avoid radiation damage to the protein) in order to firm up the SAXS result suggesting that SL causes FH 8-15 to "spring open" and lose its very compact architecture. The SPR data for FH SL and FH 8-15 SL need to be repeated several times to obtain accurate, reproducible K_D values. It would also be important to check the oligomerisation status of both FH SL and FH 3xGLY using SEC-MALS or AUC, and it would be instructive to conduct cross-linking/MS experiments to compare the architectures of wild-type and mutant FH.

Chapter 7 Summing up and future work

The current project was designed to test the hypothesis that the FH molecule adopts a dynamic bent-back conformation that allows the binding sites located close to its N and C termini to interact simultaneously with their molecular targets on a self (*i.e.* non-complement activating) surface in a selective fashion [49, 65].

For the 155-kDa FH to work in this way there must be flexibility within some parts of the protein molecule combined with a degree of stability or rigidity elsewhere. Flexibility presumably arises mainly through some of the 19 intermodular linkers, while stability could derive from interactions between some of the 20 CCP modules that are far apart from one another within the FH sequence. The central region of FH has been proposed to form a "molecular hinge"-like structure connecting more extended "arms" that embrace the target. This model was consistent with studies in our lab [52, 55] that were ongoing in the early years of the current project, which suggested that CCPs 10-15 are devoid of binding sites but form a compact structure in which non-neighbouring (in the sequence) CCPs contact one another. In the current project, work focused on a strategy of mutagenesis and structure-function studies to probe the functional anatomy of FH with a focus on the role of this putative central "hinge" in connecting up two C3b-binding sites.

First, the ability to produce sited-directed mutants of the N-terminal fragment of FH (FH 1-4), and to characterize them using functional assays and biophysical techniques, was demonstrated. Two disease-linked mutations were thus shown in the current work to be functionally deficient in differing ways.

Similar techniques were applied to mutants in the linker regions, within the contexts both of various fragments and full-length FH. The structures of fragments could be studied using NMR and SAXS while the use of EPR for characterizing such mutants was explored as a means of determining long-range distance restraints for extended and flexible proteins – such as FH and its mutants - that cannot be crystallized and are too big for NMR (see below). Functional characterization of full-length FH mutants was accomplished using standard assays while some of the shorter fragments could be tested for binding to a bacterial protein PspCN. These studies

showed for the first time that mutations in the “hinge” and in particular mutations that caused the hinge to “spring open” had an effect on the ability of Factor H to bind to C3b in a bivalent fashion.

The well-established EPR method, DEER [135, 139], was the most demanding of these techniques since it required the insertion of molecular probes into the protein at site-specific positions within the sequence. EPR along with site-directed spin labeling (SDSL) allows for distances of between 1.8 and 7 nm to be obtained, but in deuterated proteins, measurable distances can be as great as 10 nm [185]; when these distance restraints are combined with the short distances obtained from NMR the resulting structural information can address hypotheses such as the one outlined above. In the current experiments, the EPR spin labels were inserted before the first and after the last residues of the natural sequence to minimize the risk of disrupting the natural conformation of the protein. Various strategies for SDSL were attempted including the incorporation of cysteines and the engineering or proteins to include substrates for enzymatic conjugation. The successful derivation of an EPR-derived distance in FH 12-13 that matched the equivalent distances in NMR and crystal structures was a useful proof of principle. This strategy incorporated two transglutaminase tags and, in the case of the double-module construct, FH 12-13 TT, as the tag sequence is quite flexible it will be near impossible to gather exact distance measurements with this method.

Initial work presented in this thesis has thus provided a greater insight into how the structure and function of FH are related - but further work will be required.

In order to advance the study on the aHUS mutations located in CCP 1 (Chapter 3) full-length versions of FH containing these mutants will be required. A comparison of binding of C3b to these mutants, with the anticipation that the presence of the second C3b-binding site in CCPs 19-20 could influence the interaction would be performed. Also to understand how the mutant full-length proteins would affect cell-surface interactions, hemolysis assays using sheep erythrocytes coated with C3b could be used to measure both co-factor activity as well as decay acceleration activity.

Three strategies were used for labeling the proteins with the paramagnetic tags needed to gather intramolecular distances using EPR, but unfortunately all had problems and would benefit from further work (Chapter 4). Whilst the incorporation of cysteines did not appear to disrupt the natural disulphides found within FH 12-13 CC it did appear to cause problems when used in the context of FH 10-13 CC. In order to see if there is a limit to the amount of cysteines that can be present a FH fragment containing three CCPs (e.g. FH 10-12) could potentially provide some answers.

During the recording of the EPR experiments and mass spectrometry analysis of FH 12-13 CC it became apparent that only one of the labeling sites was labelled in each of the proteins. Adjusting the amounts of reducing agent, excess label and transglutaminase enzyme in the reactions may be all that is required to improve the labeling efficiency. As mentioned in Chapter 4 the use of unnatural amino acids or exploiting the N-glycosylation present in FH may be more logical routes to approach but many challenges may also arise.

In an initial experiment to compare structural changes of the FH 12-13 linker mutants it was shown that structurally there are very few changes between the three and with regards to flexibility both mutants exhibited more flexibility than the WT.

SAXS was used to monitor any structural changes that could occur in both FH 8-15 and FH 10-15 linker fragments. The FH 10-15 fragments all had a similar compact structure. For the FH 8-15 constructs it proved difficult to obtain SAXS data as the samples were extremely radiation sensitive so conditions will firstly need to be optimized to reduce the sensitivity to the x-rays prior to repeating the experiment.

The second aspect of the linker project focused on the functional changes that the modified linkers could cause (Chapter 6).

Firstly the yields of both FH 3xGLY and FH SL were poor so the use of a *P. pastoris* strain that incorporated the protein disulfide isomerise gene may help improve the yields. A further issue that needs addressing is that all the experiments were performed using samples that contained clipped protein so all of these experiments

need to be repeated using new samples but hopefully adjusting the fermentation process will reduce the amount of clipping seen.

References

1 Walport MJ. Advances in immunology: Complement (First of two parts). New England Journal of Medicine 2001, 344: 1058-1066

2 Carroll MC. The complement system in regulation of adaptive immunity. Nature Immunology 2004, 5: 981-986

3 Dunkelberger JR, Song W-C. Complement and its role in innate and adaptive immune responses. Cell Research 2010, 20: 34-50

4 Ricklin D, Hajishengallis G, Yang K, Lambris JD. Complement: a key system for immune surveillance and homeostasis. Nature Immunology 2010, 11: 785-797

5 Carroll MC. The role of complement in B cell activation and tolerance. Advances in immunology 2000, 74: 61-88

6 He JQ, Wiesmann C, Campagne MvL. A role of macrophage complement receptor CRIg in immune clearance and inflammation. Molecular Immunology 2008, 45: 4041-4047

7 Barreiro O, de la Fuente H, Mittelbrunn M, Sanchez-Madrid F. Functional insights on the polarized redistribution of leukocyte integrins and their ligands during leukocyte migration and immune interactions. Immunological Reviews 2007, 218: 147-164

8 Sahu A, Lambris JD. Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics. Immunopharmacology 2000, 49: 133-148

9 Taylor PR, Carugati A, Fadok VA, Cook HT, Andrews M, Carroll MC, Savill JS, *et al.* A hierarchical role for classical pathway complement proteins in the clearance of apoptotic cells in vivo. Journal of Experimental Medicine 2000, 192: 359-366

10 Trouw LA, Blom AM, Gasque P. Role of complement and complement regulators in the removal of apoptotic cells. Molecular Immunology 2008, 45: 1199-1207

11 Lambris JD. The multifunctional role of C3, the third component of complement. Immunology Today 1988, 9: 387-393

12 Nishida N, Walz T, Springer TA. Structural transitions of complement component C3 and its activation products. Proceedings of the National Academy of Sciences of the United States of America 2006, 103: 19737-19742

13 Janssen BJC, Huizinga EG, Raaijmakers HCA, Roos A, Daha MR, Nilsson-Ekdahl K, Nilsson B, *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 2005, 437: 505-511

14 Janssen BJC, Christodoulidou A, McCarthy A, Lambris JD, Gros P. Structure of C3b reveals conformational changes that underlie complement activity. Nature 2006, 444: 213-216

15 Wiesmann C, Katschke KJ, Jr., Yin J, Helmy KY, Steffek M, Fairbrother WJ, McCallum SA, *et al.* Structure of C3b in complex with CRIg gives insights into regulation of complement activation. Nature 2006, 444: 217-220

16 Tack BF, Harrison RA, Janatova J, Thomas ML, Prahl JW. Evidence for presence of an internal thiolester bond in third component of human complement. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 1980, 77: 5764-5768

17 Thomas ML, Janatova J, Gray WR, Tack BF. Third component of human complement: Lacalization of the internal thiolester bond. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 1982, 79: 1054-1058
18 Law SKA, Dodds AW. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. Protein Science 1997, 6: 263-274
19 Nagar B, Jones RG, Diefenbach RJ, Isenman DE, Rini JM. X-ray crystal structure of C3d: A C3 fragment and ligand for complement receptor 2. Science 1998, 280: 1277-1281
20 Dodds AW, Ren XD, Willis AC, Law SKA. The reaction mechanism of the internal thioester in the human complement component C4. Nature 1996, 379: 177-179
21 Arlaud GJ, Gaboriaud C, Thielens NM, Rossi V. Structural biology of C1. Biochemical Society Transactions 2002, 30: 1001-1006
22 Kang Y-H, Tan LA, Carroll MV, Gentle ME, Sim RB. Target Pattern Recognition by Complement Proteins of the Classical and Alternative Pathways. Target Pattern Recognition in Innate Immunity 2009, 653: 117-128
23 Duncan AR, Winter G. The binding site for C1q on IgG. Nature 1988, 332: 738-740
24 Reid KBM, Day AJ. Structure-function relationships of the complement components. Immunology Today 1989, 10:177-180
25 Dodds AW, Sim RB, Porter RR, Kerr MA. Activation of first component of human complement (C1) by antibody-antigen aggregates. Biochemical Journal 1978, 175: 383-390
26 Wallis R, Mitchell DA, Schmid R, Schwaeble WJ, Keeble AH. Paths reunited: Initiation of the classical and lectin pathways of complement activation. Immunobiology 2010, 215: 1-11
27 Thielens NM, Bersch B, Hernandez JF, Arlaud GJ. Structure and functions of the interaction domains of C1r and C1s: keystones of the architecture of the C1 complex. Immunopharmacology 1999, 42: 3-13
28 Sarma JV, Ward PA. The complement system. Cell and Tissue Research 2011, 343: 227-235
29 Takahashi K, Ezekowitz RAB. The role of the mannose-binding lectin in innate immunity. Clinical Infectious Diseases 2005, 41: 440444
30 Wallis R. Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway. Immunobiology 2007, 212: 289-299
31 Favoreel HW, Van de Walle GR, Nauwynck HJ, Pensaert MB. Virus complement evasion strategies. Journal of General Virology 2003, 84: 1-15
32 Matsushita M, Fujita T. Activation of the vlassical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. Journal of Experimental Medicine 1992, 176: 1497-502.
33 Chen CB, Wallis R. Two mechanisms for mannose-binding protein modulation of the activity of its associated serine proteases. Journal of Biological Chemistry 2004, 279: 26058-26065
34 Rawal N, Rajagopalan R, Salvi VP. Activation of complement component C5 - Comparison of C5 convertases of the lectin pathway and the classical pathway of complement. Journal of Biological Chemistry 2008, 283: 7853-7863

35 Pangburn MK, Mullereberhard HJ. Relation of a putative thioester bond in C3 to activation of the alternative pathway and the binding of C3b to biological targets of complement. Journal of Experimental Medicine 1980, 152: 1102-1114
36 Gros P, Milder FJ, Janssen BJC. Complement driven by conformational changes. Nature Reviews Immunology 2008, 8: 48-58
37 Pangburn MK, Schreiber RD, Mullereberhard HJ. C3b deposition during activation of the alternative complement pathway and the effect of deposition on the activating surface. Journal of Immunology 1983, 131: 1930-1935
38 Kirkitadze MD, Barlow PN. Structure and flexibility of the multiple domain proteins that regulate complement activation. Immunological Reviews 2001, 180: 146-161
39 Zipfel PF, Skerka C. Complement regulators and inhibitory proteins. Nature Reviews Immunology 2009, 9: 729-740
40 Nilsson UR, Mullereb.Hj. Isolation of β_{1F}-globulin from human serum and its characterization as fifth component of complement. Journal of Experimental Medicine 1965, 122: 277-298
41 Decordoba SR, Lublin DM, Rubinstein P, Atkinson JP. Human genes for 3 complement components that regulate the activation of C3 are tightly linked. Journal of Experimental Medicine 1985, 161:1189-1195
42 Ripoche J, Day AJ, Harris TJR, Sim RB. The comlete amino-acid sequence of human complement fcator H. Biochemical Journal 1988, 249: 593-602
43 de Cordoba SR, de Jorge EG. Translational mini-review series on complement factor H: Genetics and disease associations of human complement factor H. Clinical and Experimental Immunology 2008, 151:1-13
44 Hakobyan S, Harris CL, Tortajada A, De Jorge EG, Garcia-Layana A, Fernandez-Robredo P, De Cordoba SR, *et al.* Measurement of factor H variants in plasma using variant-specific monoclonal antibodies: Application to assessing risk of age-related macular degeneration. Investigative Ophthalmology & Visual Science 2008, 49: 1983-1990
45 Discipio RG. Ultrastructures and interactions of complement factor H and I. Journal of Immunology 1992, 149: 2592-2599
46 Perkins SJ, Nealis AS, Sim RB. Oligomeric domain structure of human-complement factor H by X-Ray and neutron solution scattering. Biochemistry 1991, 30: 2847-2857
47 Schmidt CQ, Herbert AP, Hocking HG, Uhrin D, Barlow PN. Translational mini-review series on complement factor H: Structural and functional correlations for factor H. Clinical and Experimental Immunology 2008, 151: 14-24
48 Morgan HP, Mertens HDT, Guariento M, Schmidt CQ, Soares DC, Svergun DI, Herbert AP, *et al.* Structural Analysis of the C-Terminal Region (Modules 18-20) of Complement Regulator Factor H (FH). Plos One 2012, 7: 1-10
49 Aslam M, Perkins SJ. Folded-back solution structure of monomeric factor H of human complement by synchrotron X-ray and neutron scattering, analytical ultracentrifugation and constrained molecular modelling. Journal of Molecular Biology 2001, 309: 1117-1138
50 Norman DG, Barlow PN, Baron M, Day AJ, Sim RB, Campbell ID. Three-dimensional structure of a complement control protein module in solution. Journal of Molecular Biology 1991, 219: 717-725

51 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009, 25: 1189-1191
52 Makou E, Matis I, Schmidt CQ, Mertens HDT, Svergun D, Barlow PN. High-resolution structures of module pairs 10-11, 11-12 and 12-13 combined with SAXS data for 10-15 and 10-18 constructs to define architecture of factor H. Molecular Immunology 2010, 47: 295-312
53 Pechtl IC, Kavanagh D, McIntosh N, Harris CL, Barlow PN. Disease-associated N-terminal Complement Factor H Mutations Perturb Cofactor and Decay-accelerating Activities. Journal of Biological Chemistry 2011, 286: 11082-11090
54 Schmidt C. Structure and function of the central part of complement factor H. PhD thesis: University of Edinburgh 2008
55 Schmidt CQ, Herbert AP, Mertens HDT, Guariento M, Soares DC, Uhrin D, Rowe AJ, *et al.* The Central Portion of Factor H (Modules 10-15) Is Compact and Contains a Structurally Deviant CCP Module. Journal of Molecular Biology 2010, 395: 105-122
56 Hocking HG, Herbert AP, Kavanagh D, Soares DC, Ferreira VP, Pangburn MK, Uhrin D, *et al.* Structure of the N-terminal region of complement factor H and conformational implications of disease-linked sequence variations. Journal of Biological Chemistry 2008, 283: 9475-9487
57 Wu J, Wu Y-Q, Ricklin D, Janssen BJC, Lambris JD, Gros P. Structure of complement fragment C3b-factor H and implications for host protection by complement regulators. Nature Immunology 2009, 10: 728-734
58 Okemefuna AI, Gilbert HE, Griggs KM, Ormsby RJ, Gordon DL, Perkins SJ. The regulatory SCR-1/5 and cell surface-binding SCR-16/20 fragments of factor H reveal partially folded-back solution structures and different self-associative properties. Journal of Molecular Biology 2008, 375: 80-101
59 Pechtl IC, Neely RK, Dryden DTF, Jones AC, Barlow PN. Use of time-resolved FRET to validate crystal structure of complement regulatory complex between C3b and factor H (N terminus). Protein Science 2011, 20: 2102-2112
60 Barlow PN, Norman DG, Steinkasserer A, Horne TJ, Pearce J, Driscoll PC, Sim RB, *et al.* Solution structure of the fifth repeat of factor H - a second example of the complement control protein module. Biochemistry 1992, 31: 3626-3634
61 Herbert AP, Deakin JA, Schmidt CQ, Blaum BS, Egan C, Ferreira VP, Pangburn MK, *et al.* Structure shows that a glycosaminoglycan and protein recognition site in factor H is perturbed by age-related macular degeneration-linked single nucleotide polymorphism. Journal of Biological Chemistry 2007, 282: 18960-18968
62 Prosser BE, Johnson S, Roversi P, Clark SJ, Tarelli E, Sim RB, Day AJ, *et al.* Expression, purification, cocrystallization and preliminary crystallographic analysis of sucrose octasulfate/human complement regulator factor H SCRs 6-8. Acta Crystallographica Section F-Structural Biology and Crystallization Communications 2007, 63: 480-483
63 Schmidt CQ, Herbert AP, Kavanagh D, Gandy C, Fenton CJ, Blaum BS, Lyon M, *et al.* A new map of glycosaminoglycan and C3b binding sites on factor H. Journal of Immunology 2008, 181: 2610-2619

64 Jokiranta TS, Jaakola V-P, Lehtinen MJ, Parepalo M, Meri S, Goldman A. Structure of complement factor H carboxyl-terminus reveals molecular basis of atypical haemolytic uremic syndrome. Embo Journal 2006, 25: 1784-1794
65 Morgan HP, Schmidt CQ, Guariento M, Blaum BS, Gillespie D, Herbert AP, Kavanagh D, *et al.* Structural basis for engagement by complement factor H of C3b on a self surface. Nature Structural & Molecular Biology 2011, 18: 463-470
66 Herbert AP, Uhrin D, Lyon M, Pangburn MK, Barlow PN. Disease-associated sequence variations congregate in a polyanion recognition patch on human factor H revealed in three-dimensional structure. Journal of Biological Chemistry 2006, 281: 16512-16520
67 Weiler JM, Daha MR, Austen KF, Fearon DT. Control of amplification convertase of complement by plasma protein β-1H. Proceedings of the National Academy of Sciences of the United States of America 1976, 73: 3268-3272
68 Gordon DL, Kaufman RM, Blackmore TK, Kwong J, Lublin DM. Identification of complement regulatory domains in human factor H. Journal of Immunology 1995, 155: 348-356
69 Jokiranta TS, Hellwage J, Koistinen V, Zipfel PF, Meri S. Each of the three binding sites on complement factor H interacts with a distinct site on C3b. Journal of Biological Chemistry 2000, 275: 27657-62.
70 Kajander T, Lehtinen MJ, Hyvaerinen S, Bhattacharjee A, Leung E, Isenman DE, Meri S, *et al.* Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. Proceedings of the National Academy of Sciences of the United States of America 2011, 108: 2897-2902
71 Ferreira VP, Herbert AP, Hocking HG, Barlow PN, Pangburn MK. Critical role of the C-terminal domains of factor H in regulating complement activation at cell surfaces. Journal of Immunology 2006, 177: 6308-6316
72 Sharma AK, Pangburn MK. Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. Proceedings of the National Academy of Sciences of the United States of America 1996, 93: 10996-11001
73 Ferreira VP, Pangburn MK, Cortes C. Complement control protein factor H: The good, the bad, and the inadequate. Molecular Immunology 2010, 47: 2187-2197
74 Fearon DT. Regulation by membrane sialic acid of β-1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. Proceedings of the National Academy of Sciences of the United States of America 1978, 75: 1971-1975
75 Pangburn MK, Atkinson MAL, Meri S. Localization of the heparin-binding site on complement factor H. Journal of Biological Chemistry 1991, 266: 16847-16853
76 Blackmore TK, Sadlon TA, Ward HM, Lublin DM, Gordon DL. Identification of a heparin binding domain in the seventh short consensus repeat of complement factor H-1. Journal of Immunology 1996, 157: 5422-5427
77 Blackmore TK, Hellwage J, Sadlon TA, Higgs N, Zipfel PF, Ward HM, Gordon DL. Identification of the second heparin-binding domain in human complement factor H. Journal of Immunology 1998, 160: 3342-3348

78 Ormsby RJ, Jokiranta TS, Duthy TG, Griggs KA, Sadlon TA, Giannakis E, Gordon DL. Localization of the third heparin-binding site in the human complement regulator factor H-1. Molecular Immunology 2006, 43: 1624-1632
79 Zipfel PF, Skerka C, Hellwage J, Jokiranta ST, Meri S, Brade V, Kraiczy P, *et al.* Factor H family proteins: on complement, microbes and human diseases. Biochemical Society Transactions 2002, 30: 971-978
80 Zipfel PF, Skerka C. FHL-1/reconectin: a human complement and immune regulator with cell-adhesive function. Immunology Today 1999, 20: 135-140
81 Skerka C, Horstmann RD, Zipfel PF. Molecular cloning of a human serum protein structurally related to complement factor H. Journal of Biological Chemistry 1991, 266: 12015-12020
82 Skerka C, Timmann C, Horstmann RD, Zipfel PF. Two additional human serum proteins structurally related to complement factor H - evidence for a family of factor H related genes. Journal of Immunology 1992, 148: 3313-3318
83 Hellwage J, Jokiranta TS, Koistinen V, Vaarala O, Meri S, Zipfel PF. Functional properties of complement factor H-related proteins FHR-3 and FHR-4: binding to the C3d region of C3b and differential regulation by heparin. Febs Letters 1999, 462: 345-352
84 Skerka C, Hellwage J, Weber W, Tilkorn A, Buck F, Marti T, Kampen E, *et al.* The human factor H-related protein 4 (FHR-4) - A novel short consensus repeat-containing protein is associated with human triglyceride-rich lipoproteins. Journal of Biological Chemistry 1997, 272: 5627-5634
85 Murphy B, Georgiou T, Machet D, Hill P, McRae J. Factor H-related protein-5: A novel component of human glomerular immune deposits. American Journal of Kidney Diseases 2002, 39: 24-27
86 Heinen S, Hartmann A, Lauer N, Wiehl U, Dahse H-M, Schirmer S, Gropp K, *et al.* Factor H-related protein 1 (CFHR-1) inhibits complement C5 convertase activity and terminal complex formation. Blood 2009, 114: 2439-2447
87 McRae JL, Cowan PJ, Power DA, Mitchelhill KI, Kemp BE, Morgan BP, Murphy BF. Human factor H-related protein 5 (FHR-5) - A new complement-associated protein. Journal of Biological Chemistry 2001, 276: 6747-6754
88 McRae JL, Duthy TG, Griggs KM, Ormsby RJ, Cowan PJ, Cromer BA, McKinstry WJ, *et al.* Human factor H-related protein 5 has cofactor activity, inhibits C3 convertase activity, binds heparin and C-reactive protein, and associates with lipoprotein. Journal of Immunology 2005, 174: 6250-6256
89 Zipfel PF, Jokiranta TS, Hellwage J, Koistinen V, Meri S. The factor H protein family. Immunopharmacology 1999, 42: 53-60
90 Jozsi M, Zipfel PF. Factor H family proteins and human diseases. Trends in Immunology 2008, 29: 389-387
91 Zipfel PF, Hallstroem T, Hammerschmidt S, Skerka C. The complement fitness Factor H: Role in human diseases and for immune escape of pathogens, like pneumococci. Vaccine 2008, 26: I67-I74
92 Lambris JD, Ricklin D, Geisbrecht BV. Complement evasion by human pathogens. Nature Reviews Microbiology 2008, 6: 132-142
93 Saunders RE, Abarrategui Garrido C, Fremeaux-Bacchi V, Goicoechea de Jorge E, Goodship THJ, Lopez Trascasa M, Noris M, *et al.* The interactive factor H-atyplical hemolytic uremic syndrome mutation database and website: Update and

integration of membrane cofactor protein and factor I mutations with structural models. Human Mutation 2007, 28: 222-234
94 Caprioli J, Castelletti F, Bucchioni S, Bettinaglio P, Bresin E, Pianetti G, Gamba S, *et al.* Complement factor H mutations and gene polymorphisms in haemolytic uraemic syndrome: the C-257T, the A2089G and the G2881T polymorphisms are strongly associated with the disease. Human Molecular Genetics 2003, 12: 3385-3395
95 Tortajada A, Montes T, Martinez-Barricarte R, Morgan BP, Harris CL, Rodriguez de Cordoba S. The disease-protective complement factor H allotypic variant Ile62 shows increased binding affinity for C3b and enhanced cofactor activity. Human Molecular Genetics 2009, 18: 3452-3461
96 Perez-Caballero D, Gonzalez-Rubio C, Gallardo ME, Vera M, Lopez-Trascasa M, de Cordoba SR, Sanchez-Corral P. Clustering of missense mutations in the C-terminal region of factor H in atypical hemolytic uremic syndrome. American Journal of Human Genetics 2001, 68: 478-484
97 Ferreira VP, Herbert AP, Cortes C, McKee KA, Blaum BS, Esswein ST, Uhrin D, *et al.* The Binding of Factor H to a Complex of Physiological Polyanions and C3b on Cells Is Impaired in Atypical Hemolytic Uremic Syndrome. Journal of Immunology 2009, 182: 7009-7018
98 Noris M, Caprioli J, Bresin E, Mossali C, Pianetti G, Gamba S, Daina E, *et al.* Relative Role of Genetic Complement Abnormalities in Sporadic and Familial aHUS and Their Impact on Clinical Phenotype. Clinical Journal of the American Society of Nephrology 2010, 5: 1844-1859
99 Kavanagh D, Goodship THJ. Atypical hemolytic uremic syndrome, genetic basis, and clinical manifestations. Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program 2011, 2011
100 Appel GB, Cook HT, Hageman G, Jennette JC, Kashgarian M, Kirschfink M, Lambris JD, *et al.* Membranoproliferative glomerulonephritis type II (dense deposit disease): An update. Journal of the American Society of Nephrology 2005, 16: 1392-1403
101 Licht C, Heinen S, Jozsi M, Loeschmann I, Saunders RE, Perkins SJ, Waldherr R, *et al.* Deletion of Lys224 in regulatory domain 4 of Factor H reveals a novel pathomechanism for dense deposit disease (MPGN II). Kidney International 2006, 70: 42-50
102 Khandhadia S, Cipriani V, Yates JRW, Lotery AJ. Age-related macular degeneration and the complement system. Immunobiology 2012, 217: 127-146
103 Troutbeck R, Al-Qureshi S, Guymer RH. Therapeutic targeting of the complement system in age-related macular degeneration: a review. Clinical and Experimental Ophthalmology 2012, 40: 18-26
104 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, *et al.* Complement factor H polymorphism in age-related macular degeneration. Science 2005, 308: 385-389
105 Chamberlain M, Baird P, Dirani M, Guymer R. Unraveling a complex genetic disease: Age-related macular degeneration. Survey of Ophthalmology 2006, 51: 576-586

106 Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. Proceedings of the National Academy of Sciences of the United States of America 2005, 102:7227-7237
107 Thakkinstian A, Han P, McEvoy M, Smith W, Hoh J, Magnusson K, Zhang K, *et al.* Systematic review and meta-analysis of the association between complementary factor HY402H polymorphisms and age-related macular degeneration. Human Molecular Genetics 2006, 15: 242-246
108 Clark SJ, Bishop PN, Day AJ. Complement factor H and age-related macular degeneration: the role of glycosaminoglycan recognition in disease pathology. Biochemical Society Transactions 2010, 38: 1342-1348
109 Zipfel PF, Wuerzner R, Skerka C. Complement evasion of pathogens: Common strategies are shared by diverse organisms. Molecular Immunology 2007, 44: 3850-3857
110 Kraiczy P, Wurzner R. Complement escape of human pathogenic bacteria by acquisition of complement regulators. Molecular Immunology 2006, 43: 31-44
111 Jarva H, Jokiranta TS, Wurzner R, Meri S. Complement resistance mechanisms of streptococci. Molecular Immunology 2003, 40: 95-107
112 Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. Nature Reviews Microbiology 2008, 6: 288-301
113 Hammerschmidt S. Adherence molecules of pathogenic pneumococci. Current Opinion in Microbiology 2006, 9: 12-20
114 Brooks-Walter A, Briles DE, Hollingshead SK. The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive antibodies to PspA and provides immunity to pneumococcal bacteremia. Infection and Immunity 1999, 67: 6533-6542
115 Iannelli F, Oggioni MR, Pozzi G. Allelic variation in the highly polymorphic locus pspC of Streptococcus pneumoniae. Gene 2002, 284: 63-71
116 Luo RS, Mann B, Lewis WS, Rowe A, Heath R, Stewart ML, Hamburger AE, *et al.* Solution structure of choline binding protein A, the major adhesin of Streptococcus pneumoniae. Embo Journal 2005, 24: 34-43
117 Hammerschmidt S, Agarwal V, Kunert A, Haelbich S, Skerka C, Zipfel PF. The host immune regulator factor H interacts via two contact sites with the PspC protein of Streptococcus pneamoniae and mediates adhesion to host epithelial cells. Journal of Immunology 2007, 178: 5848-5858
118 Agarwal V, Asmat TM, Luo S, Jensch I, Zipfel PF, Hammerschmidt S. Complement Regulator Factor H Mediates a Two-step Uptake of Streptococcus pneumoniae by Human Cells. Journal of Biological Chemistry 2010, 285: 23484-23493
119 Barnes JP, Liang ZC, McHaourab HS, Freed JH, Hubbell WL. A multifrequency electron spin resonance study of T4 lysozyme dynamics. Biophysical Journal 1999, 76: 3298-3306
120 Weil JA, Bolton JR. Electron Paramagnetis Resonance Elemetery Theory and Practical Applications. Second Edition: John Wiley & Sons, Inc. Hoboken, New Jersey 2007

121 Schiemann O, Prisner TF. Long-range distance determinations in biomacromolecules by EPR spectroscopy. Quarterly Reviews of Biophysics 2007, 40: 1-53
122 Hubbell WL, Cafiso DS, Altenbach C. Identifying conformational changes with site-directed spin labeling. Nature Structural Biology 2000, 7: 735-739
123 Fanucci GE, Cafiso DS. Recent advances and applications of site-directed spin labeling. Current Opinion in Structural Biology 2006, 16: 644-653
124 Hubbell WL, McHaourab HS, Altenbach C, Lietzow MA. Watching proteins move using site-directed spin labeling. Structure 1996, 4: 779-783
125 Xie JM, Schultz PG. Adding amino acids to the genetic repertoire. Current Opinion in Chemical Biology 2005, 9: 548-554
126 Dougherty DA. Unnatural amino acids as probes of protein structure and function. Current Opinion in Chemical Biology 2000, 4: 645-652
127 Sato H, Ikeda M, Suzuki K, Hirayama K. Site-specific modification of interleukin-2 by the combined use of genetic engineering techniques and transglutaminase. Biochemistry 1996, 35: 13072-13080
128 Lundblad RL. Chemical Reagents for Protein Modification. Third ed.:CRC Press 2005
129 Folk JE. Transglutaminases. Annual Review of Biochemistry 1980, 49: 517-531
130 Dutton A, Singer SJ. Crosslinking and labeling of membrane proteins by transglutaminase-catalyzed reactions. Proceedings of the National Academy of Sciences of the United States of America 1975, 72: 2568-2571
131 Gorman JJ, Folk JE. Structural features of glutamine substrates for transglutaminases - role of extended interactions in the specificity of human-plasma factor-XIIIA and of the guinea-pig liver enzyme. Journal of Biological Chemistry 1984, 259: 9007-9010
132 Coussons PJ, Price NC, Kelly SM, Smith B, Sawyer L. Transglutaminase catalyzes the modification of glutamine side chains in the C-terminal region of bovine beta-lactoglobulin. Biochemical Journal 1992, 283: 803-806
133 Taki M, Shiota M, Taira K. Transglutaminase-mediated N- and C-terminal fluorescein labeling of a protein can support the native activity of the modified protein. Protein Engineering Design & Selection 2004, 17: 119-126
134 Altenbach C, Kusnetzow AK, Ernst OP, Hofmann KP, Hubbell WL. High-resolution distance mapping in rhodopsin reveals the pattern of helix movement due to activation. Proceedings of the National Academy of Sciences of the United States of America 2008, 105: 7439-7444
135 Steinhoff HJ, Savitsky A, Wegener C, Pfeiffer M, Plato M, Mobius K. High-field EPR studies of the structure and conformational changes of site-directed spin labeled bacteriorhodopsin. Biochimica Et Biophysica Acta-Bioenergetics 2000, 1457: 253-262
136 Klare JP, Steinhoff H-J. Spin labeling EPR. Photosynthesis Research 2009, 102: 377-390
137 Gaffney BJ, Bradshaw MD, Frausto SD, Wu F, Freed JH, Borbat P. Locating a Lipid at the Portal to the Lipoxygenase Active Site. Biophysical Journal 2012, 103: 2134-2144

138 Hilger D, Polyhach Y, Padan E, Jung H, Jeschke G. High-resolution structure of a Na+/H+ antiporter dimer obtained by pulsed election paramagnetic resonance distance measurements. Biophysical Journal 2007, 93: 3675-3683
139 McHaourab HS, Steed PR, Kazmier K. Toward the Fourth Dimension of Membrane Protein Structure: Insight into Dynamics from Spin-Labeling EPR Spectroscopy. Structure 2011, 19: 1549-1561
140 Schmidt CQ, Slingsby FC, Richards A, Barlow PN. Production of biologically active complement factor H in therapeutically useful quantities. Protein Expression and Purification 2011, 76: 254-263
141 Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. Fems Microbiology Reviews 2000, 24: 45-66
142 Macauley-Patrick S, Fazenda ML, McNeil B, Harvey LM. Heterologous protein production using the Pichia pastoris expression system. Yeast 2005, 22: 249-270
143 Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. Molecular and Cellular Biology 1989, 9: 1316-1323
144 Cregg JM, Vedvick TS, Raschke WC. Recent advances in the expression of foreign genes in Pichia pastoris. Bio-Technology 1993, 11: 905-910
145 Gellissen G. Heterologous protein production in methylotrophic yeasts. Applied Microbiology and Biotechnology 2000, 54: 741-750
146 Bodenhausen G, Ruben DJ. Natural abundance N15 NMR by enhanced heteronuclear spectroscopy. Chemical Physics Letters 1980, 69: 185-189
147 Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPIPE - A multidimensional spectral processing system based on UNIX pipes. Journal of Biomolecular Nmr 1995, 6: 277-293
148 Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas P, Ulrich EL, *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins-Structure Function and Bioinformatics 2005, 59: 687-696
149 Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. Journal of the American Chemical Society 1982, 104: 4546-4559
150 Clore GM, Szabo A, Bax A, Kay LE, Driscoll PC, Gronenborn AM. Deviations from the simple 2-parameter model-free approach to the interpretation of N15 nuclear magnetic relaxation of proteins. Journal of the American Chemical Society 1990, 112: 4989-4991
151 Sahu A, Isaacs SN, Soulika AM, Lambris JD. Interaction of vaccinia virus complement control protein with human complement proteins: Factor I-mediated degradation of C3b to iC3b(1) inactivates the alternative complement pathway. Journal of Immunology 1998, 160: 5596-5604
152 Pangburn MK, Mullereberhard HJ. Kinetics and thermodynamic analysis of the control of C3b by the complement regulatory proteins factor H and factor I. Biochemistry 1983, 22: 178-185
153 Provaznikova D, Rittich S, Malina M, Seeman T, Marinov I, Riedl M, Hrachovinova I. Manifestation of atypical hemolytic uremic syndrome caused by novel mutations in MCP. Pediatric Nephrology 2012, 27: 73-81

154 Heurich M, Martinez-Barricarte R, Francis NJ, Roberts DL, Rodriguez de Cordoba S, Morgan BP, Harris CL. Common polymorphisms in C3, factor B, and factor H collaborate to determine systemic complement activity and disease risk. Proceedings of the National Academy of Sciences of the United States of America 2011, 108: 8761-8766
155 Roversi P, Johnson S, Caesar JJE, McLean F, Leath KJ, Tsiftsoglou SA, Morgan BP, *et al.* Structural basis for complement factor I control and its disease-associated sequence polymorphisms. Proceedings of the National Academy of Sciences of the United States of America 2011, 108: 12839-12844
156 Lehtinen MJ, Rops AL, Isenman DE, van der Vlag J, Jokiranta TS. Mutations of Factor H Impair Regulation of Surface-bound C3b by Three Mechanisms in Atypical Hemolytic Uremic Syndrome. Journal of Biological Chemistry 2009, 284: 15650-15658
157 Kavanagh D, Anderson HE. Interpretation of genetic variants of uncertain significance in atypical hemolytic uremic syndrome. Kidney Int 2012, 81: 11-13
158 Heinen S, Jozsi M, Hartmann A, Noris M, Remuzzi G, Skerka C, Zipfel PF. Hemolytic uremic syndrome: A factor H mutation (E1172Stop) causes defective complement control at the surface of endothelial cells. Journal of the American Society of Nephrology 2007, 18: 506-514
159 Nan R, Gor J, Perkins SJ. Implications of the progressive self-association of wild-type human factor H6 for complement regulation and disease. Journal of Molecular Biology 2008, 375: 891-900
160 Pangburn MK, Rawal N, Cortes C, Alam MN, Ferreira VP, Atkinson MAL. Polyanion-Induced Self-Association of Complement Factor H. Journal of Immunology 2009, 182: 1061-1068
161 Okemefuna AI, Li K, Nan R, Ormsby RJ, Sadlon T, Gordon DL, Perkins SJ. Multimeric Interactions between Complement Factor H and Its C3d Ligand Provide New Insight on Complement Regulation. Journal of Molecular Biology 2009, 391: 119-135
162 Pechtl IC. A study of complement regulatory factor H based on Forster resonance energy transfer and investigation of disease-linked genetic variants. PhD thesis. University of Edinburgh 2010
163 Kwon T, Biebuyck N, Cailliez M, Broux F, Macher MA, Niaudet P, Tsimaratos M, *et al.* Plasmatherapy in Atypical Hemolytic Uremic Syndrome, French Cohort. Pediatric Nephrology 2010, 25: 1827-1827
164 Kavanagh D, Goodship T. Genetics and complement in atypical HUS. Pediatric Nephrology 2010, 25: 2431-2442
165 Byeon I-JL, Meng X, Jung J, Zhao G, Yang R, Ahn J, Shi J, *et al.* Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function. Cell 2009, 139: 780-790
166 Lovett JE, Lovett BW, Harmer J. DEER-Stitch: Combining three- and four-pulse DEER measurements for high sensitivity, deadtime free data. Journal of Magnetic Resonance 2012, 223: 98-106
167 Schmidt C. Structure and function of the central part of complement factor H. PhD thesis. University of Edinburgh 2008

168 Jeschke G, Chechik V, Ionita P, Godt A, Zimmermann H, Banham J, Timmel CR, *et al.* DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. Applied Magnetic Resonance 2006, 30: 473-498
169 Chiang YW, Borbat PP, Freed JH. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. Journal of Magnetic Resonance 2005, 172: 279-295
170 Czogalla A, Pieciul A, Jezierski A, Sikorski AF. Attaching a spin to a protein - site-directed spin labeling in structural biology. Acta Biochimica Polonica 2007, 54: 235-244
171 Longhi S, Belle V, Fournel A, Guigliarelli B, Carriere F. Probing structural transitions in both structured and disordered proteins using site-directed spin-labeling EPR spectroscopy. Journal of Peptide Science 2011, 17: 315-328
172 Voss J, Salwinski L, Kaback HR, Hubbell WL. A method for distance determination in proteins using a designed metal ion binding site and site-directed spin labeling: Evaluation with T4 lysozyme. Proceedings of the National Academy of Sciences of the United States of America 1995, 92: 12295-12299
173 Ritter DW, Roberts JR, McShane MJ. Glycosylation site-targeted PEGylation of glucose oxidase retains native enzymatic activity. Enzyme and Microbial Technology 2013, 52: 279-285
174 Veronese FM, Pasut G. PEGylation, successful approach to drug delivery. Drug Discovery Today 2005, 10: 1451-1458
175 Atanassov II, Atanassov II, Etchells JP, Turner SR. A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. Plant Methods 2009, 5
176 Zhang H, Loovers HM, Xu L-Q, Wang M, Rowling PJE, Itzhaki LS, Gong W, *et al.* Alcohol oxidase (AOX1) from Pichia pastoris is a novel inhibitor of prion propagation and a potential ATPase. Molecular Microbiology 2009, 71: 702-716
177 Clore GM, Gronenborn AM, Bax A. A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. Journal of Magnetic Resonance 1998, 133: 216-221
178 Woessner DE. Nuclear spin relaxation in ellipsoids undergoing rotational brownian motion. Journal of Chemical Physics 1962, 37: 647-654
179 Petoukhov MV, Svergun DI. Applications of small-angle X-ray scattering to biomacromolecular solutions. The international journal of biochemistry & cell biology 2013, 45: 429-437
180 Mertens HDT, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. Journal of Structural Biology 2010, 172: 128-141
181 Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI. Structural characterization of flexible proteins using small-angle X-ray scattering. Journal of the American Chemical Society 2007, 129: 5656-5664
182 Franke D, Svergun DI. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography 2009, 42: 342-346

183 Lu L, Ma YY, Zhang JR. Streptococcus pneumoniae recruits complement factor H through the amino terminus yof CbpA. Journal of Biological Chemistry 2006, 281: 15464-15474
184 Makou E. The architecture of the central region of factor H and its interaction with PspC of *S. pneumoniae*. PhD thesis,University of Edinburgh 2013
185 Jeschke G. DEER Distance Measurements on Proteins. Annual Review of Physical Chemistry, Vol 63 2012, 63: 419-446

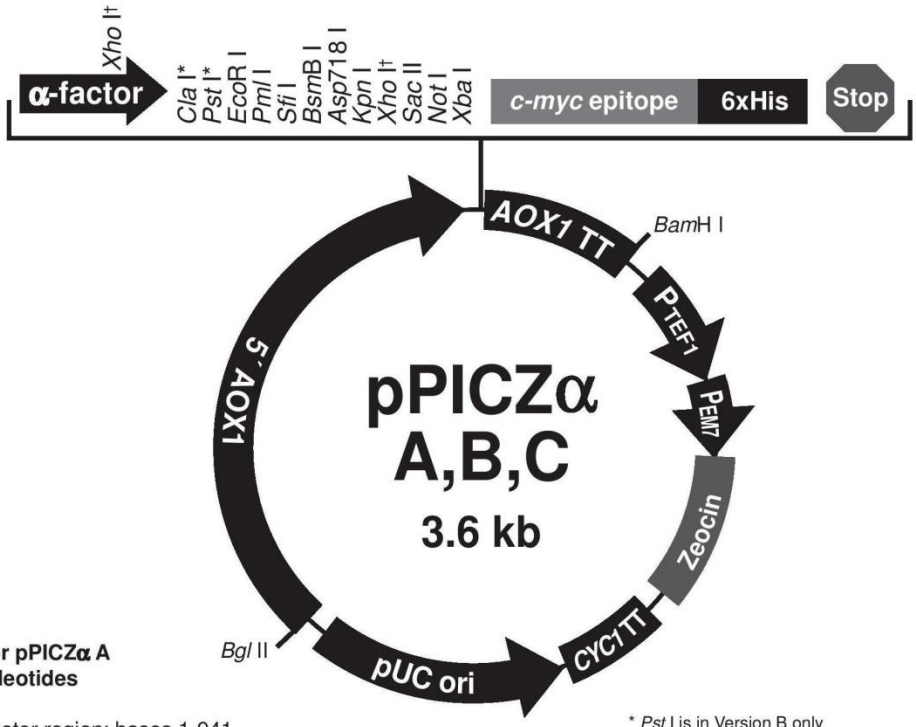
Appendices

Appendix A Media and buffers

MEDIA	COMPOSITION (w/v unless stated otherwise)
Basal Salts for 15N-labelled fermentor growth	4% (v/v) glycerol 0.413% potassium hydroxide 1.5% magnesium sulphate heptahydrate 1.82% potassium sulphate 0.093% calcium sulphate 2.67% (v/v) ortho-phosphoric acid
BMGY (Buffered glycerol-complex)	100 mM potassium phosphate pH 6 1.34 % YNB $4x10^{-5}$ % biotin 1 % glycerol 1% yeast extract 2% peptone
BMMY (Buffered methanol complex)	100 mM potassium phosphate pH 6 1.34 % YNB $4x10^{-5}$ % biotin 0.5 % methanol 1% yeast extract 2% peptone

Low-salt LB (Luria-Bertani) (low salt is needed for use with zeocin antibiotic)	0.5 % yeast extract 1 % tryptone 0.5 % sodium chloride +/- 1.5 % agar
LB (Luria-Bertani)	0.5 % yeast extract 1 % tryptone 1 % sodium chloride +/- 1.5 % agar
YPD (Yeast extract, Peptone, Dextrose)	1 % yeast extract 2 % peptone 2 % dextrose (D-glucose) +/- 1.5 % agar
EDTA stock solution (ethylenediaminetetraacetic acid)	0.5 M stock adjusted to pH with sodium hydroxide
TAE buffer 50x in distilled H_2O	Tris-acetate (2 M) EDTA (100 mM)
TAE Agarose	Agarose 1 % in TAE buffer

Appendix B vector map and sequencing primers



Comments for pPICZα A
3593 nucleotides

5´ *AOX1* promoter region: bases 1-941
5´ *AOX1* priming site: bases 855-875
α-factor signal sequence: bases 941-1207
Multiple cloning site: bases 1208-1276
c-myc epitope: bases 1275-1304
Polyhistidine (6xHis) tag: bases 1320-1337
3´ *AOX1* priming site: bases 1423-1443
AOX1 transcription termination region: bases 1341-1682
TEF1 promoter: bases 1683-2093
EM7 promoter: bases 2095-2162
Sh ble ORF: bases 2163-2537
CYC1 transcription termination region: bases 2538-2855
pUC origin: bases 2866-3539 (complementary strand)

* *Pst* I is in Version B only
Cla I is in Version C only

†The two *Xho* I sites in the vector allow the user to clone their gene in frame with the Kex2 cleavage site, resulting in expression of their native gene without additional amino acids at the N-terminus.

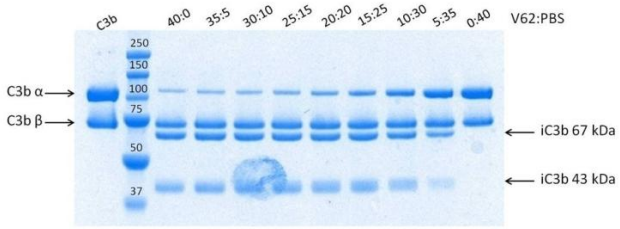
Sequencing primers:

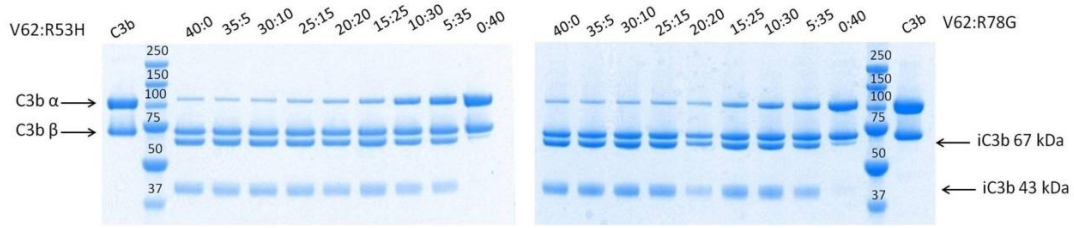
Name	Sequence (5’ – 3’)
5’ Sequencing	GGGGATTTCGATGTTGCTGTTTTG
3’ Sequencing	CCGGTCTTCTCGTAAGTGCC

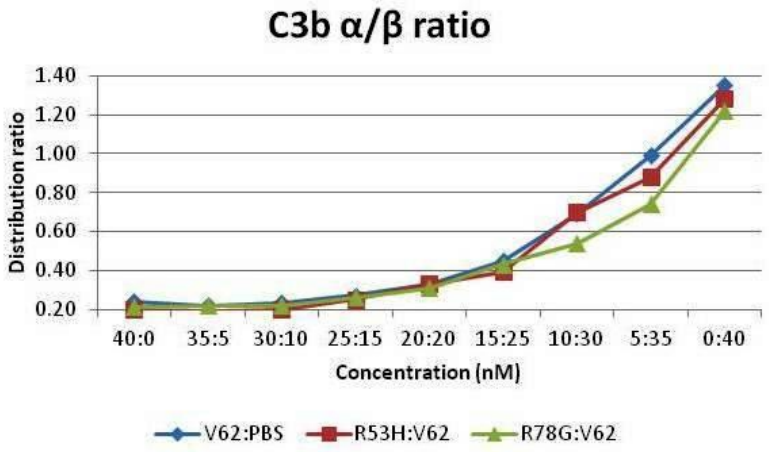
Appendix C Quantification of C3b α and β bands using distribution ratios

As an additional way to quantify the inhibition assay seen in section 3.2.2 was the use of distribution ratios. Ratios were calculated using the equation W=X/Y where X was the densitometry of the C3b α band and Y the densitometry of the C3b β band.

For all three experiments the ratios were very similar between 40:0 and 25:15 followed by a slight increase for the 20:20 and 15:25 samples then an approximate linear relationship was seen between 10:30 and 0:40. As the results are very similar for each of the experiments it can be concluded that the disease linked mutants R53H and R78G do not disrupt the cofactor ability of "wild type" factor H.







Appendix D Protein Summary

In the table below all the constructs are summarized including all the non-native sequences found at the N-terminus or C-terminus as well as the FH residue numbers that they consist of. The purification method for the proteins used is shown below.

CONSTRUCT	FACTOR H RESIDUE NUMBERS	NON-NATIVE N TERMINUS	NON-NATIVE C TERMINUS
1-4 I62	19-263	AGEQKLISEEDL	HHHHHH
1-4 R53H	19-263	AGEQKLISEEDL	HHHHHH
1-4 R78G	19-263	AGEQKLISEEDL	HHHHHH
11-14 CT	628-866	ACG	PKPQQFM
10-15CT	568-927	ACG	PKPQQFM
CODOPT 10-13 WT	567-804	AG	-
CODOPT 10-13 CC	567-804	AGC	C
CODOPT 10-13 TT	567-804	AGPKPQQFM	PKPQQFM
CODOPT 10-13 CT	567-804	AGC	PKPQQFM
CODOPT 10-13 TC	567-804	AGPKPQQFM	C
CODOPT 12-13 WT	690-804	AG	-
CODOPT 12-13 CC	690-804	AGC	C
CODOPT 12-13 TT	690-804	AGPKPQQFM	PKPQQFM
CODOPT 12-13 CT	690-804	AGC	PKPQQFM
CODOPT 12-13 TC	690-804	AGPKPQQFM	C
CODOPT 12-13 3xGLY	690-804	AG	-
12-13 SHORT LINKER	690-804	AG	-
CODOPT 8-15 WT	447-927	AG	-
CODOPT 8-15 3xGLY	447-927	AG	-

CODOPT 8-15 SHORT LINKER	447-927	AG	-
CODOPT 10-15	568-927	AG	-
CODOPT 10-15 3xGLY	568-927	AG	-
CODOPT 10-15 SHORT LINKER	568-927	AG	-
CODOPT FH 3xGLY	1-1217	AG	-
CODOPT FH SHORT LINKER	1-1217	AG	-

Sequences of FH constructs. Numbering includes the 18 amino acid residues that form the natural signal peptide found at the start of FH (UniProtKB entry ID P08603).

CONSTRUCT	STARTING MATERIAL	FLASK/ FERMENTOR	ENDO HF	CATCHMENT STEP	PURIFICATION STEP 1	PURIFICATION STEP 2	SEQUENCE VALIDATION	FOLD VALIDATION	COMMENTS
1-4 I62	1-4 I62 DNA in pPICZαB	FLASK	NO	NICKEL AFFINITY	SIZE EXCLUSION			COFACTOR ASSAY	
1-4 R53H	1-4 R53H DNA in pPICZαB	FLASK	NO	NICKEL AFFINITY	SIZE EXCLUSION			COFACTOR ASSAY	LOW YIELD
1-4 R78G	1-4 R78G DNA in pPICZαB	FLASK	NO	NICKEL AFFINITY	SIZE EXCLUSION			COFACTOR ASSAY	LOW YIELD
10-15CT	10-15 DNA in pPICZαB.	FLASK	YES		SIZE EXCLUSION	SIZE EXCLUSION	Trypsin Digest	1D 1H NMR	
11-14 CT	11-14 DNA in pPICZαB.	FLASK	YES	SP SEPHAROSE	SIZE EXCLUSION	-	Trypsin Digest	1D 1H NMR	
12-13 SHORT LINKER (SL)	12-13 SHORT LINKER DNA in pPICZαB	FERMENTOR	YES	SP SEPHAROSE	SIZE EXCLUSION	SIZE EXCLUSION		1D 1H NMR ^{15}N HSQC	
CODOPT 12-13 WT	FULL LENGTH FH DNA	FERMENTOR	YES	SP SEPHAROSE	SIZE EXCLUSION	SIZE EXCLUSION		1D 1H NMR ^{15}N HSQC	
CODOPT 12-13 CC	FULL LENGTH FH DNA	FLASK	YES	SP SEPHAROSE	SIZE EXCLUSION		Accurate MS	1D 1H NMR	ONLY 1 CYS FREE
CODOPT 12-13 TT	FULL LENGTH FH DNA	FLASK	YES	SP SEPHAROSE					

CODOPT 12-13 3xGLY	FULL LENGTH FH DNA	FERMENTOR	YES	SP SEPHAROSE	SIZE EXCLUSION	SIZE EXCLUSION		1D 1H NMR ^{15}N HSQC	
CODOPT 8-15 WT	FULL LENGTH FH DNA	FLASK	YES	Q SEPHAROSE	SIZE EXCLUSION	MONO Q	Trypsin Digest		
CODOPT 8-15 3xGLY	FULL LENGTH FH DNA	FERMENTOR	YES	SP SEPHAROSE	MONO Q	SIZE EXCLUSION	Trypsin Digest		
CODOPT 8-15 SL	FULL LENGTH FH DNA	FERMENTOR	YES	SP SEPHAROSE	MONO Q	SIZE EXCLUSION	Trypsin Digest		
CODOPT 10-15 WT	FULL LENGTH FH DNA	FLASK	YES	SP SEPHAROSE	MONO Q	SIZE EXCLUSION	Trypsin Digest		
CODOPT 10-15 3xGLY	FULL LENGTH FH DNA	FERMENTOR	YES	SP SEPHAROSE	MONO Q	SIZE EXCLUSION	Trypsin Digest		
CODOPT 10-15 SL	FULL LENGTH FH DNA	FERMENTOR	YES	SP SEPHAROSE	MONO Q	SIZE EXCLUSION	Trypsin Digest		
CODOPT FH 3xGLY	FULL LENGTH FH DNA	FERMENTOR	NO	SP SEPHAROSE	SIZE EXCLUSION			COFACTOR ASSAY	
CODOPT FH SL	GENEART	FERMENTOR	NO	SP SEPHAROSE	SIZE EXCLUSION			COFACTOR ASSAY	